

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2001, 16:06:38 ; Search time 26.49 Seconds

(without alignments)  
51.633 Million cell updates/sec

Title: US-08-908-867-39

Perfect score: 121  
Sequence: 1 XXXGTXXXXSKOXEEAVRLXXXLLKNGXSGAXXXXX 40

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

A.Geneseq\_36:\*

1: /SIDSI/gcgdata/geneseq/geneseqp/AA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseqp/AA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseqp/AA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseqp/AA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseqp/AA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseqp/AA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseqp/AA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseqp/AA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseqp/AA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseqp/AA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseqp/AA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseqp/AA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseqp/AA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseqp/AA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseqp/AA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseqp/AA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseqp/AA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseqp/AA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseqp/AA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseqp/AA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseqp/AA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	94	77.7	37	20	Y17618
2	93	76.9	36	20	Y17606
3	93	76.9	37	20	Y24869
4	93	76.9	37	20	Y24853
5	93	76.9	37	20	Y24854
6	92	76.0	35	20	Y31535
7	92	76.0	35	20	Y24839
8	92	76.0	35	20	Y17608
9	92	76.0	36	20	Y31533
10	92	76.0	36	20	Y24856
11	92	76.0	36	20	Y24837
12	92	76.0	36	20	Y17619

13	92	76.0	36	20	Y17605	Exendin agonist pe
14	92	76.0	37	20	Y31531	Exendin agonist pe
15	92	76.0	37	20	Y31549	Exendin agonist pe
16	92	76.0	37	20	Y31550	Exendin agonist pe
17	92	76.0	37	20	Y31565	Exendin agonist pe
18	92	76.0	37	20	Y24855	Exendin agonist pe
19	92	76.0	37	20	Y24835	Exendin agonist pe
20	92	76.0	38	20	Y31529	Exendin agonist pe
21	92	76.0	38	20	Y31547	Exendin agonist pe
22	92	76.0	38	20	Y24851	Exendin agonist pe
23	92	76.0	38	20	Y24833	Exendin agonist pe
24	92	76.0	38	20	Y17616	Exendin agonist pe
25	92	76.0	38	20	Y17603	Exendin agonist pe
26	92	76.0	39	19	W61773	Leu(14), Phe(25)-e
27	92	76.0	39	20	Y17623	Exendin agonist pe
28	91	75.2	35	20	Y31534	Exendin agonist pe
29	91	75.2	35	20	Y31553	Exendin agonist pe
30	91	75.2	35	20	Y24857	Exendin agonist pe
31	91	75.2	35	20	Y24838	Exendin agonist pe
32	91	75.2	35	20	Y17620	Exendin agonist pe
33	91	75.2	35	20	Y17607	Exendin agonist pe
34	91	75.2	35	21	Y78961	Exendin-4 (1-35)
35	91	75.2	36	20	Y31532	Exendin agonist pe
36	91	75.2	36	20	Y31552	Exendin agonist pe
37	91	75.2	36	20	Y24836	Exendin agonist pe
38	91	75.2	36	21	Y78960	Exendin-4 (1-36)
39	91	75.2	37	20	Y31530	Exendin agonist pe
40	91	75.2	37	20	Y31551	Exendin agonist pe
41	91	75.2	37	21	Y78959	Exendin-4 (1-37)
42	91	75.2	38	20	Y31528	Exendin agonist pe
43	91	75.2	38	20	Y31548	Exendin agonist pe
44	91	75.2	38	20	Y24852	Exendin agonist pe
45	91	75.2	38	20	Y24832	Exendin agonist pe

#### ALIGNMENTS

RESULT 1

ID Y17618 standard; peptide: 37 AA.

AC Y17618;

XX 09-AUG-1999 (first entry)

DE Exendin agonist peptide #84.

XX

XX Exendin; agonist; Heloderma sp.; Gila monster; venom; lizard;

KW diabetes mellitus type I; diabetes mellitus type II; hyperglycaemia;

KW hypoglycaemia; plasma glucose; gastric emptying; stomach emptying.

XX

OS Synthetic.

OS Heloderma sp.

XX

PN W09925728-A1.

XX

PD 27-MAY-1999.

XX

PF 13-NOV-1998; 98WO-US24273.

XX

PR 14-NOV-1997; 97US-0066029.

XX

PA (AMYL-) AMYLIN PHARM INC.

XX	Beeley NRA, Prickett KS;
XX	WPI; 1999-347456/29.
XX	Peptide agonists of exendin - delay stomach emptying, for treating
PT	diabetes and hypo- or hyper-glycaemia
XX	Claim 28; Fig 4; 144p; English.

XX Y17535 to Y17624 represent extendin peptide agonists. Extendins are  
 CC peptides that are found in the venom of the Gila-monster, a lizard  
 CC endogenous to Arizona and Northern Mexico. The peptide agonists are  
 CC used to treat diabetes mellitus (types I or II), hyperglycaemia or  
 CC hypoglycaemia. They can also be used for in vitro and in vivo studies  
 CC on extendins and their agonists. They regulate gastric motility and slow  
 CC gastric emptying (resulting in lower post-prandial glucose levels).  
 XX  
 SQ Sequence 37 AA:

Query Match 77.7%; Score 94; DB 20; Length 37;  
 Best Local Similarity 65.6%; Pred. No. 1.2e-10;  
 Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 4 GTXXXXXSKQXEEAVRLXXXXLKNGXSSGA 35  
 DB 4 gfttsalskqmeeeavrlflewlnkgsssga 35

RESULT 2  
 Y17606  
 ID Y17606 standard; peptide: 36 AA.

XX Y17606;

DT 09-AUG-1999 (first entry)

XX Extendin agonist peptide #72.

DE Extendin; agonist; Heloderma sp.; Gila monster; venom; lizard;

XX diabetes mellitus type I; diabetes mellitus type II; hyperglycaemia;

XX hypoglycaemia; plasma glucose; gastric emptying; stomach emptying.

XX Synthetic.

OS Heloderma sp.

XX W09925728-A1.

PD 27-MAY-1999.

PF 13-NOV-1998; 98WO-US24273.

XX 14-NOV-1997; 97US-0066029.

PA (AMYL-) AMTILIN PHARM INC.

XX Beeley NRA, Prickett KS;

PI WPI; 1999-347456/29.

DR Peptide agonists of extendin - delay stomach emptying, for treating

XX diabetes and hypo- or hyper-glycaemia

PS Claim 28; Fig 4; 144pp; English.

XX Y17535 to Y17624 represent extendin peptide agonists. Extendins are  
 CC peptides that are found in the venom of the Gila-monster, a lizard  
 CC endogenous to Arizona and Northern Mexico. The peptide agonists are  
 CC used to treat diabetes mellitus (types I or II), hyperglycaemia or  
 CC hypoglycaemia. They can also be used for in vitro and in vivo studies  
 CC on extendins and their agonists. They regulate gastric motility and slow  
 CC gastric emptying (resulting in lower post-prandial glucose levels).  
 XX

SQ Sequence 36 AA:

Query Match 76.9%; Score 93; DB 20; Length 36;  
 Best Local Similarity 65.6%; Pred. No. 1.8e-10;  
 Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 4 GTXXXXXSKQXEEAVRLXXXXLKNGXSSGA 35

DB 4 gfttsalskqmeeeavrlflewlnkgsssga 35

RESULT 3

Y24869  
 ID Y24869 standard; peptide: 37 AA.

XX Y24869;

DT 24-AUG-1999 (first entry)

XX Extendin agonist peptide #61.

DE Extendin; agonist; Heloderma sp.; Gila monster; venom; lizard;

XX diabetes mellitus type I; diabetes mellitus type II; hyperglycaemia;

XX hypoglycaemia; plasma glucose; gastric emptying; stomach emptying.

XX Synthetic.

OS Heloderma sp.

XX W09925727-A2.

PD 27-MAY-1999.

PF 13-NOV-1998; 98WO-US24210.

XX 14-NOV-1997; 97US-0065442.

PA (AMYL-) AMTILIN PHARM INC.

XX Beeley NRA, Prickett KS;

PI WPI; 1999-394773/33.

DR New extendin agonist peptides - can regulate gastric motility and

XX slow gastric emptying, used for treating, e.g. diabetes

XX Claim 18; Fig 4; 108pp; English.

CC Y24809 to Y24877 represent extendin agonist peptides which can regulate  
 CC gastric motility and slow gastric emptying. The peptides can be used for  
 CC treating e.g. diabetes or hyperglycaemic or hypoglycaemic conditions.

CC The peptides are extendin agonists which have activity as agents to  
 CC regulate gastric motility and to slow gastric emptying, as evidenced by  
 CC the ability to reduce post-prandial glucose levels in mammals. They can  
 CC be used for the treatment of type I and II diabetes and hyperglycaemic  
 CC or hypoglycaemic conditions. They can also be used for the treatment of  
 CC disorders which would be benefited by agents which lower plasma glucose  
 CC levels and in treatment of disorders which would be benefited with  
 CC agents useful in delaying and/or slowing gastric emptying.

XX

SQ Sequence 37 AA:

Query Match 76.9%; Score 93; DB 20; Length 37;  
 Best Local Similarity 68.8%; Pred. No. 1.8e-10;  
 Matches 22; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

OY 4 GTXXXXXSKQXEEAVRLXXXXLKNGXSSGA 35  
 DB 4 gfttsalskqmeeeavrlflewlnkgsssga 35

RESULT 4

Y24853  
 ID Y24853 standard; peptide: 37 AA.

XX Y24853;

DT 24-AUG-1999 (first entry)

DE Extendin agonist peptide #45.

XX	Exendin; agonist; Heloderma sp.; Gila monster; venom; lizard.
KW	diabetes mellitus type I; diabetes mellitus type II; hyperglycaemia;
KW	hypoglycaemia; plasma glucose; gastric emptying; stomach emptying.
XX	
OS	Synthetic.
OS	Heloderma sp.
XX	
XX	WO9925727-A2.
PN	
XX	
PD	27-MAY-1999.
XX	
PF	13-NOV-1998; 98WO-US24210.
XX	
PR	14-NOV-1997; 97US-0065442.
XX	
PA	(AMYL-) AMYLIN PHARM INC.
XX	
PI	Beeley NRA, Prickett KS;
XX	
DR	WPI; 1999-394773/33.
XX	
PT	New exendin agonist peptides - can regulate gastric motility and
PT	slow gastric emptying, used for treating, e.g. diabetes
XX	
PS	Claim 18; Fig 4; 108bp; English.
XX	
CC	Y24809 to Y24877 represent exendin agonist peptides which can regulate
CC	gastric motility and slow gastric emptying. The peptides can be used for
CC	treating e.g. diabetes or hyperglycaemic or hypoglycaemic conditions.
CC	The peptides are exendin agonists which have activity as agents to
CC	regulate gastric motility and to slow gastric emptying, as evidenced by
CC	the ability to reduce post-prandial glucose levels in mammals. They can
CC	be used for the treatment of Type I and II diabetes and hyperglycaemic
CC	or hypoglycaemic conditions. They can also be used for the treatment of
CC	disorders which would be benefited by agents which lower plasma glucose
CC	levels and in treatment of disorders which would be benefited with
CC	agents useful in delaying and/or slowing gastric emptying.
XX	
SO	Sequence 37 AA;
XX	
Query Match	76.9%; Score 93; DB 20; Length 37;
Best Local Similarity	65.6%; Pred. No. 1.8e-10;
Matches 21; Conservative	0; Mismatches 11; Indels 0; Gaps 0;
Oy	4 GTFXXXXKXQXEAEAVRLXXXXLKNKGXSSGA 35
Db	4 gftfcdiskqmeeavrlflewknkggassa 35
XX	
RESULT	5
Y24854	
Y24854	standard; peptide; 37 AA.
XX	
AC	Y24854;
XX	
DT	24-AUG-1999 (first entry)
XX	
DE	Exendin agonist peptide #46.
XX	
KW	Exendin; agonist; Heloderma sp.; Gila monster; venom; lizard;
KW	diabetes mellitus type I; diabetes mellitus type II; hyperglycaemia;
XX	hypoglycaemia; plasma glucose; gastric emptying; stomach emptying.
XX	
OS	Synthetic.
OS	Heloderma sp.
XX	
PN	WO9925727-A2.
XX	
PD	27-MAY-1999.
XX	
PF	13-NOV-1998; 98WO-US24210.

XX	14-NOV-1997;	97US-0065442.
PR		
XX		
PA	(AMYL-) AMYLIN PHARM INC.	
XX		
PI	Beeley NRA, Prickett KS;	
XX		
DR	WPI; 1999-394773/33.	
XX		
PT	New extendin agonist peptides - can regulate gastric motility and	
PT	slow gastric emptying, used for treating, e.g. diabetes	
XX		
PS	Claim 18; Fig 4; 108pp; English.	
XX		
CC	Y24809 to Y24877 represent extendin agonist peptides which can regulate	
CC	gastric motility and slow gastric emptying. The peptides can be used for	
CC	treating e.g. diabetes or hyperglycaemic or hypoglycaemic conditions.	
CC	The peptides are extendin agonists which have activity as agents to	
CC	regulate gastric motility and to slow gastric emptying, as evidenced by	
CC	the ability to reduce post-prandial glucose levels in mammals. They can	
CC	be used for the treatment of Type I and II diabetes and hyperglycaemic	
CC	or hypoglycaemic conditions. They can also be used for the treatment of	
CC	disorders which would be benefited by agents which lower plasma glucose	
CC	levels and in treatment of disorders which would be benefited with	
CC	agents useful in delaying and/or slowing gastric emptying.	
XX		
SQ	Sequence 37 AA;	
	Query Match 76.9%; Score 93; DB 20; Length 37;	
	Best Local Similarity 65.6%; Pred. No. 1.8e-10;	
	Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0	
QY	4 GTXXXXXKQXEEAVRLXXXXLKNKGXSCGA 35	
Db	4 gtfstdskqmeeeavrlfiewlknkgasga 35	
RESULT 6		
Y31535		
ID	Y31535 standard; peptide: 35 AA.	
XX		
AC	Y31535;	
XX		
DT	08-NOV-1999 (first entry)	
XX		
DE	Extendin agonist peptide.	
XX		
KW	Extendin; agonist; GLP-1; glucagon-like peptide; toxic hypervolemia;	
KW	diuresis; renal plasma flow; glomerular filtration rate; pre-eclampsia;	
KW	eclampsia of pregnancy; cardiac contractility; renal failure; diuretic;	
KW	congestive heart failure; nephrotic syndrome; pulmonary edema; cirrhosis;	
KW	hypertension; urine flow.	
XX		
OS	Synthetic.	
OS	Heloderma sp.	
XX		
FT	Key Location/Qualifiers	
FT	Modified-site 35 /note="C-terminal amide"	
XX		
PN	WO9940788-A1.	
XX		
PD	19-AUG-1999.	
XX		
PF	05-FEB-1999; 99WO-US02554.	
XX		
PR	13-FEB-1998; 98US-0075122.	
XX		
PA	(AMYL-) AMYLIN PHARM INC.	
PI	Beeley NRA, Prickett K, Vine W, Young AA;	
XX		

DR WPI; 1999-527332/44.  
 XX Increasing urine flow by administering peptides or peptide agonists  
 XX  
 XX Example 34; Page 48; 94pp; English.  
 PS  
 XX The invention relates to new methods of increasing urine flow that  
 CC comprises administering an extendin or extendin agonist, or a GLP-1  
 CC (glucagon-like peptide) or GLP-1 agonist. The new methods using an  
 CC extendin, extendin agonist, GLP-1 or GLP-1 agonist are useful for  
 CC increasing urine flow, decreasing potassium concentration in urine,  
 CC preventing or alleviating a disorder associated with toxic hypervolemia  
 CC (renal failure, congestive heart failure, nephrotic syndrome, pulmonary  
 CC edema, cirrhosis, or hypertension). They can also be used for inducing  
 CC rapid diuresis, preparing an individual for surgical procedure,  
 CC increasing renal plasma flow and glomerular filtration rate, treating  
 CC pre-eclampsia or eclampsia of pregnancy, and increasing a condition/  
 CC disorder that can be alleviated by increasing cardiac contractility  
 CC (congestive heart failure, pulmonary edema, systemic edema or renal  
 CC failure). Unlike prior art diuretics, the new methods increase urine  
 CC excretion and sodium excretion without increasing potassium loss, and are  
 CC fast acting. They have a prolonged duration of action, are inotropic,  
 CC have a low toxicity, and are easily administered intravenously.  
 CC Sequences Y31505-560 represent examples of extendin agonists compounds.  
 CC  
 XX  
 SQ Sequence 35 AA;  
 Query Match 76.0%; Score 92; DB 20; Length 35;  
 Best Local Similarity 65.6%; Pred. No. 2.6e-10;  
 Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
 QY 4 GTYXXXSKQEEAVRLXXXLKNKGXSSGA 35  
 Db 4 gftfslskqleeeavrlfiefllkngpssga 35  
 RESULT 7  
 Y24839  
 ID Y24839 standard; peptide; 35 AA.  
 AC Y24839;  
 XX  
 XX 24-AUG-1999 (first entry)  
 DT XX  
 DE Extendin agonist peptide #31.  
 XX  
 XX Extendin; agonist; Heloderma sp.; Gila monster; venom; lizard;  
 KW diabetes mellitus type I; diabetes mellitus type II; hyperglycaemia;  
 KW hypoglycaemia; plasma glucose; gastric emptying; stomach emptying.  
 XX  
 OS Synthetic.  
 OS Heloderma sp.  
 XX  
 PN WO925727-A2.  
 XX  
 PD 27-MAY-1999.  
 PF 13-NOV-1998; 98WO-US24210.  
 XX  
 PR 14-NOV-1997; 97US-0065442.  
 XX  
 PA (AMYL-) AMYLIN PHARM INC.  
 XX  
 PI Bealey NRA, Prickett KS;  
 XX  
 DR WPI; 1999-394773/33.  
 XX  
 PT New extendin agonist peptides - can regulate gastric motility and  
 PT slow gastric emptying, used for treating, e.g. diabetes  
 XX  
 PS Claim 18; Fig 4; 108pp; English.  
 XX

CC Y24809 to Y24877 represent extendin agonist peptides which can regulate  
 CC gastric motility and slow gastric emptying. The peptides can be used for  
 CC treating e.g. diabetes or hyperglycaemic or hypoglycaemic conditions.  
 CC The peptides are extendin agonists which have activity as agents to  
 CC regulate gastric motility and to slow gastric emptying, as evidenced by  
 CC the ability to reduce post-prandial glucose levels in mammals. They can  
 CC be used for the treatment of type I and II diabetes and hyperglycaemic  
 CC or hypoglycaemic conditions. They can also be used for the treatment of  
 CC disorders which would be benefited by agents which lower plasma glucose  
 CC levels and in treatment of disorders which would be benefited with  
 CC agents useful in delaying and/or slowing gastric emptying.  
 CC  
 XX  
 SQ Sequence 35 AA;  
 Query Match 76.0%; Score 92; DB 20; Length 35;  
 Best Local Similarity 65.6%; Pred. No. 2.6e-10;  
 Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
 QY 4 GTYXXXSKQEEAVRLXXXLKNKGXSSGA 35  
 Db 4 gftfslskqleeeavrlfiefllkngpssga 35  
 RESULT 8  
 Y17608  
 ID Y17608 standard; peptide; 35 AA.  
 AC Y17608;  
 XX  
 XX 09-AUG-1999 (first entry)  
 DT XX  
 DE Extendin agonist peptide #74.  
 XX  
 XX Extendin; agonist; Heloderma sp.; Gila monster; venom; lizard;  
 KW diabetes mellitus type I; diabetes mellitus type II; hyperglycaemia;  
 KW hypoglycaemia; plasma glucose; gastric emptying; stomach emptying.  
 XX  
 OS Synthetic.  
 OS Heloderma sp.  
 XX  
 PN WO925728-A1.  
 XX  
 PD 27-MAY-1999.  
 PF 13-NOV-1998; 98WO-US24273.  
 XX  
 PR 14-NOV-1997; 97US-0066029.  
 XX  
 PA (AMYL-) AMYLIN PHARM INC.  
 XX  
 PI Bealey NRA, Prickett KS;  
 XX  
 DR WPI; 1999-347456/29.  
 XX  
 PT Peptide agonists of extendin - delay stomach emptying, for treating  
 PT diabetes and hypo- or hyper-glycaemia  
 XX  
 PS Claim 28; Fig 4; 144pp; English.  
 XX  
 XX Y17335 to Y17624 represent extendin peptide agonists. Extendins are  
 CC peptides that are found in the venom of the Gila-monster, a lizard  
 CC endogenous to Arizona and Northern Mexico. The peptide agonists are  
 CC used to treat diabetes mellitus (types I or II), hyperglycaemia or  
 CC hypoglycaemia. They can also be used for in vitro and in vivo studies  
 CC on extendins and their agonists. They regulate gastric motility and slow  
 CC gastric emptying (resulting in lower post-prandial glucose levels).  
 CC  
 XX  
 SQ Sequence 35 AA;  
 Query Match 76.0%; Score 92; DB 20; Length 35;  
 Best Local Similarity 65.6%; Pred. No. 2.6e-10;  
 Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 4 GTXXXXXSKC~~EE~~EAVRLXXXXLKN~~G~~SGA 35  
|| ||.||||| |||| ||||  
Db 4 gtftsdlskqleeeavrlfiefllkn~~g~~ppssga 35

RESULT	10
Y24856	
ID	Y24856 standard; peptide; 36 AA
XX	

AC Y24856;  
XX

DT 24-AUG-1999 (first entry)  
XX

DE      Exendin agonist peptide #48  
XX

OS Synthetic.  
OS Heloderma sp.  
KW Extendin; agonist: Heloderma sp.: Gila monster; venom; lizard;  
KW diabetes; agonist: Heloderma sp.: Gila monster; venom; lizard;  
KW diabetes mellitus type I; diabetes mellitus type II; hyperglycaemia  
XX hypoglycaemia; plasma glucose; gastric emptying; stomach emptying.

XX  
PN W09925727-A2

PN W09925727-A2  
XX

PD 27-MAY-1999.  
XX  
DT 13-MAY-1999.

PF	13-NOV-1998;	98WO-US24210
XX		
14	NOV 1998	0773 005745

PR 14-NOV-1997; 97US-0065442  
XX

PA (AMYL-) AMYLIN PHARM INC.  
XX

PI Beeley NRA, Prickett  
XX  
1000 204550 000

DR WPI; 1999-394773/33  
XX

PT New extendin agonist peptides - can regulate gastric motility and slow gastric emptying, used for treating, e.g. diabetes

XX	Claim 18; Fig 4; 108pp; English
PS	
W	

XX Y24809 to Y24877 represent extendin agonist peptides which can regulate  
CC extendin receptor 1 expression in the developing brain.

gastric motility and slow gastric emptying. The peptides can be used for treating e.g. diabetes or hyperglycaemic or hypoglycaemic conditions.

CC The peptides are extendin agonists which have activity as agents to regulate gastric motility and to slow gastric emptying, as evidenced by

the ability to reduce post-prandial glucose levels in mammals. They can be used for the treatment of Type I and II diabetes and hyperglycaemic

or hypoglycaemic conditions. They can also be used for the treatment of disorders which would be benefited by agents which lower plasma glucose

CC levels and in treatment of disorders which would be benefited with CC agents useful in delaying and/or slowing gastric emptying.

XX	Sequence	36 AA;
SQ		

100

Query Match	76.0%;	Score 92;	DB 20;	Length 36;
Best Local Similarity	68.8%;	Pred. No. 2.7e-10;		

Matches	22;	Conservative	0;	Mismatches	10;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	-----	--------	----	------	----

QY 4 GTXXXXXSKXEEXEAVRLXXXXLKNKGXSSGA 35

Db 4 gtttsdlskqmeearvrlfiewlknqgxssga 35

RESULT	11
Y24837	

AC Y24837;  
XX  
DT 24-AUG-1999 (first entry)

```

XX DE Extendin agonist peptide #29.
XX XX
XX XX Extendin; agonist; Heloderma sp.; Gila monster; venom; lizard;
XX KW diabetes mellitus type I; diabetes mellitus type II; hyperglycaemia;
XX KW hypoglycaemia; plasma glucose; gastric emptying; stomach emptying.
XX OS Synthetic.
XX OS Heloderma sp.
XX PN WO9925727-A2.
XX PD 27-MAY-1999.
XX PF 13-NOV-1998; 98WO-US24210.
XX PR 14-NOV-1997; 97US-0065442.
XX PA (AMYL-) AMTLIN PHARM INC.
XX PI Bealey NRA, Prickett KS;
XX DR WPI; 1999-34773/33.
XX PT New extendin agonist peptides - can regulate gastric motility and
XX PT slow gastric emptying, used for treating, e.g. diabetes
XX PS Claim 18; Fig 4; 108pp; English.
XX CC Y24809 to Y24877 represent extendin agonist peptides which can regulate
XX CC gastric motility and slow gastric emptying. The peptides can be used for
XX CC treating e.g. diabetes or hyperglycaemic or hypoglycaemic conditions.
XX CC The peptides are extendin agonists which have activity as agents to
XX CC regulate gastric motility and to slow gastric emptying, as evidenced by
XX CC the ability to reduce post-prandial glucose levels in mammals. They can
XX CC be used for the treatment of Type I and II diabetes and hyperglycaemic
XX CC or hypoglycaemic conditions. They can also be used for the treatment of
XX CC disorders which would be benefited by agents which lower plasma glucose
XX CC levels and in treatment of disorders which would be benefited with
XX CC agents useful in delaying and/or slowing gastric emptying.
XX SQ Sequence 36 AA;

Query Match 76.0%; Score 92; DB 20; Length 36;
Best Local Similarity 65.6%; Pred. No. 2.7e-10;
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 4 GTXXXXSKQXEEAVRLXXXXLKNKGXSSGA 35
   ||| ||| ||| ||| ||| ||| |||
Db 4 gftsdlskqleeeavrllflefllknkgssga 35

RESULT 12
Y17619 Y17619 standard; peptide; 36 AA.
XX
XX AC Y17619;
XX
XX DT 09-AUG-1999 (first entry)
XX
XX DE Extendin agonist peptide #85.
XX
XX KW Extendin; agonist; Heloderma sp.; Gila monster; venom; lizard;
XX KW diabetes mellitus type I; diabetes mellitus type II; hyperglycaemia;
XX KW hypoglycaemia; plasma glucose; gastric emptying; stomach emptying.
XX OS Synthetic.
XX OS Heloderma sp.
XX PN WO9925728-A1.
XX PD 27-MAY-1999.

```

```

XX XX
XX PF 13-NOV-1998; 98WO-US24273.
XX
XX PR 14-NOV-1997; 97US-0066029.
XX
XX PA (AMYL-) AMTLIN PHARM INC.
XX
XX PI Bealey NRA, Prickett KS;
XX DR WPI; 1999-347456/29.
XX
XX PT Peptide agonists of extendin - delay stomach emptying, for treating
XX PT diabetes and hypo- or hyper-glycaemia
XX PS Claim 28; Fig 4; 144pp; English.
XX CC Y17535 to Y17624 represent extendin peptide agonists. Extendins are
XX CC peptides that are found in the venom of the Gila-monster, a lizard
XX CC endogenous to Arizona and Northern Mexico. The peptide agonists are
XX CC used to treat diabetes mellitus (types I or II), hyperglycaemia or
XX CC hypoglycaemia. They can also be used for in vitro and in vivo studies
XX CC on extendins and their agonists. They regulate gastric motility and slow
XX CC gastric emptying (resulting in lower post-prandial glucose levels).
XX SQ Sequence 36 AA;

Query Match 76.0%; Score 92; DB 20; Length 36;
Best Local Similarity 68.8%; Pred. No. 2.7e-10;
Matches 22; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 4 GTXXXXSKQXEEAVRLXXXXLKNKGXSSGA 35
   ||| ||| ||| ||| ||| ||| |||
Db 4 gftsdlskqmeeavrllflefllknkgssga 35

RESULT 13
Y17605 Y17605 standard; peptide; 36 AA.
XX
XX AC Y17605;
XX
XX DT 09-AUG-1999 (first entry)
XX
XX DE Extendin agonist peptide #71.
XX
XX KW Extendin; agonist; Heloderma sp.; Gila monster; venom; lizard;
XX KW diabetes mellitus type I; diabetes mellitus type II; hyperglycaemia;
XX KW hypoglycaemia; plasma glucose; gastric emptying; stomach emptying.
XX OS Synthetic.
XX OS Heloderma sp.
XX PN WO9925728-A1.
XX
XX PD 27-MAY-1999.
XX
XX PF 13-NOV-1998; 98WO-US24273.
XX
XX PR 14-NOV-1997; 97US-0066029.
XX
XX PA (AMYL-) AMTLIN PHARM INC.
XX
XX PI Bealey NRA, Prickett KS;
XX DR WPI; 1999-347456/29.
XX
XX PT Peptide agonists of extendin - delay stomach emptying, for treating
XX PT diabetes and hypo- or hyper-glycaemia
XX PS Claim 28; Fig 4; 144pp; English.
XX CC Y17535 to Y17624 represent extendin peptide agonists. Extendins are

```

CC peptides that are found in the venom of the Gila-monster, a lizard  
 CC endogenous to Arizona and Northern Mexico. The peptide agonists are  
 CC used to treat diabetes mellitus (types I or II), hyperglycaemia or  
 CC hypoglycaemia. They can also be used for in vitro and in vivo studies  
 CC on exendins and their agonists. They regulate gastric motility and slow  
 CC gastric emptying (resulting in lower post-prandial glucose levels).

SQ Sequence 36 AA;

Query Match 76.0%; Score 92; DB 20; Length 36;  
 Best Local Similarity 65.6%; Pred. No. 2.7e-10;  
 Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 4 GTXXXXXSKQEEAVRLXXXXLKNGXSSGA 35  
 || ||| ||||| ||||| |||||  
 Db 4 gtttsalskqmeeeavrlfiewlknngpsssa 35

RESULT 14  
 Y31531  
 ID X31531 standard; peptide; 37 AA.

AC X31531;  
 DT 08-NOV-1999 (first entry)

DE Exendin agonist peptide.

KW Exendin; agonist; GLP-1; glucagon-like peptide; toxic hypervolemia;  
 KW diuresis; renal plasma flow; glomerular filtration rate; pre-eclampsia;  
 KW eclampsia of pregnancy; cardiac contractility; renal failure; diuretic;  
 KW congestive heart failure; nephrotic syndrome; pulmonary edema; cirrhosis;  
 KW hypertension; urine flow.

OS Synthetic.  
 OS Heloderma sp.

Key Location/Qualifiers  
 FT Modified-site 38 /note="C-terminal amide"

PN W09940788-A1.

PD 19-AUG-1999.

PF 05-FEB-1999; 99WO-US02554.

PR 13-FEB-1998; 98US-0075122.

PA (AMYL-) AMYLIN PHARM INC.

PI Beeley NRA, Prickett K, Vine W, Young AA;

DR WPI; 1999-527332/44.

PT Increasing urine flow by administering peptides or peptide agonists

PS Example 30; Page 46; 94pp; English.

CC The invention relates to new methods of increasing urine flow that  
 CC comprises administering an exendin or exendin agonist, or a GLP-1  
 CC (glucagon-like peptide) or GLP-1 agonist. The new methods using an  
 CC exendin, exendin agonist, GLP-1 or GLP-1 agonist are useful for  
 CC increasing urine flow, decreasing potassium concentration in urine,  
 CC preventing or alleviating a disorder associated with toxic hypervolemia  
 CC (renal failure, congestive heart failure, nephrotic syndrome, pulmonary  
 CC edema, cirrhosis, or hypertension). They can also be used for inducing  
 CC rapid diuresis, preparing an individual for surgical procedure,  
 CC increasing renal plasma flow and glomerular filtration rate, treating  
 CC pre-eclampsia or eclampsia of pregnancy, and increasing a condition/  
 CC disorder that can be alleviated by increasing cardiac contractility  
 CC (congestive heart failure, pulmonary edema, systemic edema or renal

CC failure). Unlike prior art diuretics, the new methods increase urine  
 CC excretion and sodium excretion without increasing potassium loss, and are  
 CC fast acting. They have a prolonged duration of action, are isotropic,  
 CC have a low toxicity, and are easily administered intravenously.

CC Sequences Y31505-560 represent examples of exendin agonists compounds.

SQ Sequence 37 AA;

Query Match 76.0%; Score 92; DB 20; Length 37;  
 Best Local Similarity 65.6%; Pred. No. 2.8e-10;  
 Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 4 GTXXXXXSKQEEAVRLXXXXLKNGXSSGA 35  
 || ||| ||||| ||||| |||||  
 Db 4 gtttsalskqleeeavrlfiefikngpsssa 35

RESULT 15  
 Y31549  
 ID X31549 standard; peptide; 37 AA.

AC X31549;

DT 08-NOV-1999 (first entry)

DE Exendin agonist peptide.

KW Exendin; agonist; GLP-1; glucagon-like peptide; toxic hypervolemia;  
 KW diuresis; renal plasma flow; glomerular filtration rate; pre-eclampsia;  
 KW eclampsia of pregnancy; cardiac contractility; renal failure; diuretic;  
 KW congestive heart failure; nephrotic syndrome; pulmonary edema; cirrhosis;  
 KW hypertension; urine flow.

OS Synthetic.  
 OS Heloderma sp.

Key Location/Qualifiers  
 FT Modified-site 31 /note="N-methyl alanine"  
 FT Modified-site 37 /note="C-terminal amide"

PN W09940788-A1.

PD 19-AUG-1999.

PF 05-FEB-1999; 99WO-US02554.

PR 13-FEB-1998; 98US-0075122.

PA (AMYL-) AMYLIN PHARM INC.

PI Beeley NRA, Prickett K, Vine W, Young AA;

DR WPI; 1999-527332/44.

PT Increasing urine flow by administering peptides or peptide agonists

PS Example 48; Page 55; 94pp; English.

CC The invention relates to new methods of increasing urine flow that  
 CC comprises administering an exendin or exendin agonist, or a GLP-1  
 CC (glucagon-like peptide) or GLP-1 agonist. The new methods using an  
 CC exendin, exendin agonist, GLP-1 or GLP-1 agonist are useful for  
 CC increasing urine flow, decreasing potassium concentration in urine,  
 CC preventing or alleviating a disorder associated with toxic hypervolemia  
 CC (renal failure, congestive heart failure, nephrotic syndrome, pulmonary  
 CC edema, cirrhosis, or hypertension). They can also be used for inducing  
 CC rapid diuresis, preparing an individual for surgical procedure,  
 CC increasing renal plasma flow and glomerular filtration rate, treating  
 CC pre-eclampsia or eclampsia of pregnancy, and increasing a condition/  
 CC disorder that can be alleviated by increasing cardiac contractility





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2001, 16:07:46 : Search time 22.07 Seconds  
(without alignments)  
32.546 Million cell updates/sec

Title: US-08-908-867-39

Perfect score: 121  
Sequence: 1 XXXGTXXXXXXKQEEAVRLXXXXXKNGXSGAXXXXX 40

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/PCRTUS.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/Backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	91	75.2	39	1	US-08-066-480-1 Sequence 1, Appl
2	91	75.2	39	1	US-08-066-480-2 Sequence 2, Appl
3	83	68.6	31	1	US-08-066-480-3 Sequence 5, Appl
4	75	62.0	31	1	US-08-066-480-4 Sequence 3, Appl
5	75	62.0	31	1	US-08-066-480-4 Sequence 4, Appl
6	41	33.9	589	2	US-08-317-305-2 Sequence 4, Appl
7	41	33.9	589	2	US-08-317-305-4 Sequence 4, Appl
8	41	33.9	589	3	US-08-862-508-2 Sequence 2, Appl
9	41	33.9	589	3	US-08-862-508-4 Sequence 4, Appl
10	41	33.9	589	4	PCT-US95-12508-2 Sequence 4, Appl
11	41	33.9	589	4	PCT-US95-12508-4 Sequence 4, Appl
12	39	32.2	341	1	US-08-062-024B-5 Sequence 5, Appl
13	39	32.2	341	1	US-08-891-254-5 Sequence 5, Appl
14	39	32.2	341	2	US-08-756-407-5 Sequence 5, Appl
15	39	32.2	341	2	US-08-819-539-5 Sequence 5, Appl
16	39	32.2	341	2	US-09-030-270A-5 Sequence 5, Appl
17	39	32.2	341	4	PCT-US94-05014-5 Sequence 5, Appl
18	39	32.2	341	4	PCT-US96-08819-5 Sequence 5, Appl
19	38.5	31.8	651	2	US-08-492-027A-1 Sequence 1, Appl
20	38.5	31.8	655	2	US-08-492-027A-6 Sequence 6, Appl
21	38	31.4	357	1	US-08-552-142A-2 Sequence 2, Appl
22	38	31.4	357	1	US-08-910-973-2 Sequence 2, Appl
23	38	31.4	357	4	PCT-US95-05741-2 Sequence 2, Appl
24	36	29.8	318	1	US-08-220-958-4 Sequence 4, Appl
25	35	28.9	156	1	US-08-552-142A-9 Sequence 9, Appl
26	35	28.9	156	1	US-08-910-973-9 Sequence 9, Appl
27	35	28.9	156	4	PCT-US95-05741-9 Sequence 9, Appl
28	35	28.9	356	1	US-08-552-142A-15 Sequence 15, Appl

29	35	28.9	356	1	US-08-910-973-15	Sequence 15, Appl
30	35	28.9	356	3	US-09-234-332-7	Sequence 7, Appl
31	35	28.9	356	3	US-09-234-332-8	Sequence 8, Appl
32	35	28.9	546	2	US-08-492-027A-8	Sequence 8, Appl
33	35	28.9	619	3	US-08-813-150-6	Sequence 6, Appl
34	35	28.9	777	2	US-08-231-193A-16	Sequence 16, Appl
35	35	28.9	777	2	US-08-486-273A-16	Sequence 16, Appl
36	35	28.9	777	2	US-08-486-273A-16	Sequence 16, Appl
37	35	28.9	777	3	US-08-940-086A-16	Sequence 16, Appl
38	35	28.9	854	2	US-08-231-193A-32	Sequence 32, Appl
39	35	28.9	854	2	US-08-486-273A-32	Sequence 32, Appl
40	35	28.9	854	3	US-08-480-474-32	Sequence 32, Appl
41	35	28.9	854	3	US-08-940-086A-32	Sequence 32, Appl
42	35	28.9	863	3	US-08-436-332B-2	Sequence 2, Appl
43	35	28.9	870	2	US-08-231-193A-30	Sequence 30, Appl
44	35	28.9	870	2	US-08-486-273A-30	Sequence 30, Appl
45	35	28.9	870	3	US-08-480-474-30	Sequence 30, Appl

#### ALIGNMENTS

```
RESULT 1
US-08-066-480-1
; Sequence 1, Application US/08066480
; Patent No. 5424286
; GENERAL INFORMATION:
; APPLICANT: Eng, John
; TITLE OF INVENTION: Pharmaceutical Compositions And Use of
; TITLE OF INVENTION: Exendin-3 and Exendin-4 for Treatment of Diabetes Mellitus
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Allegretti & Wilcoff, Ltd.
; STREET: 10 S. Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/066,480
; FILING DATE: 24-MAR-1993
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: McDonnell, John J
; REGISTRATION NUMBER: 26,949
; REFERENCE/DOCKET NUMBER: 93,084
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-715-1000
; TELEFAX: 312-715-1234
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 39 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..39
; OTHER INFORMATION: /label= Exendin-3
US-08-066-480-1
```

Query Match 75.2% Score 91: DB 1: Length 39:  
Best Local Similarity 65.6% Pred. No. 5.9e-10:  
Matches 21: Conservative 0: Mismatches 11: Indels 0: Gaps 0:  
QY 4 GTXXXXXKQEEAVRLXXXXXKNGXSGAXSGA 35

Db 4 GTFTSLSKQMEEAVALRLEMLKNGPSSGA 35

RESULT 2  
US-08-066-480-2  
Sequence 2, Application US/08066480  
Patent No. 5424286

## GENERAL INFORMATION:

APPLICANT: Eng, John  
TITLE OF INVENTION: Pharmaceutical Compositions And Use of  
TITLE OF INVENTION: Exendin-3 and Exendin-4 for Treatment of Diabetes Mellitus  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Allegretti & Witcoff, Ltd.  
STREET: 10 S. Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/066,480  
FILING DATE: 24-MAR-1993

## CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:  
NAME: McDonnell, John J  
REGISTRATION NUMBER: 26,949  
REFERENCE/DOCKET NUMBER: 93,084  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-715-1000  
TELEFAX: 312-715-1234

## INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:  
LENGTH: 39 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..39

OTHER INFORMATION: /label= Exendin-4

US-08-066-480-2

Query Match 75.2%; Score 91; DB 1; Length 39;  
Best Local Similarity 65.6%; Pred. No. 5.9e-10;  
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 4 GTXXXXXKQXEEAVALRXXXXLXKNGXSSGA 35  
Db 4 GTFTSLSKQMEEAVALRLEMLKNGPSSGA 35

RESULT 3  
US-08-066-480-5

Sequence 5, Application US/08066480  
Patent No. 5424286

## GENERAL INFORMATION:

APPLICANT: Eng, John  
TITLE OF INVENTION: Pharmaceutical Compositions And Use of  
TITLE OF INVENTION: Exendin-3 and Exendin-4 for Treatment of Diabetes Mellitus  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Allegretti & Witcoff, Ltd.  
STREET: 10 S. Wacker Drive  
CITY: Chicago  
STATE: Illinois

COUNTRY: USA

ZIP: 60606

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/066,480  
FILING DATE: 24-MAR-1993  
CLASSIFICATION: 514

## ATTORNEY/AGENT INFORMATION:

NAME: McDonnell, John J  
REGISTRATION NUMBER: 26,949  
REFERENCE/DOCKET NUMBER: 93,084  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-715-1000  
TELEFAX: 312-715-1234

## INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:  
LENGTH: 31 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:

NAME/KEY: Peptide

LOCATION: 1..31

OTHER INFORMATION: /label= Exendin-9-39

US-08-066-480-5

Query Match 68.6%; Score 83; DB 1; Length 31;  
Best Local Similarity 76.0%; Pred. No. 1.3e-08;

Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 11 SKQXEEAVALRXXXXLXKNGXSSGA 35  
Db 3 SKQMEEAVALRLEMLKNGPSSGA 27

RESULT 4  
US-08-066-480-3

Sequence 3, Application US/08066480  
Patent No. 5424286

## GENERAL INFORMATION:

APPLICANT: Eng, John  
TITLE OF INVENTION: Pharmaceutical Compositions And Use of  
TITLE OF INVENTION: Exendin-3 and Exendin-4 for Treatment of Diabetes Mellitus  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Allegretti & Witcoff, Ltd.  
STREET: 10 S. Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/066,480  
FILING DATE: 24-MAR-1993

## CLASSIFICATION: 514

## ATTORNEY/AGENT INFORMATION:

NAME: McDonnell, John J  
REGISTRATION NUMBER: 26,949  
REFERENCE/DOCKET NUMBER: 93,084  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-715-1000

TELEFAX: 312-715-1234  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 31 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..31  
OTHER INFORMATION: /label= Extendin-1-31  
US-08-066-480-3

Query Match 62.0%; Score 75; DB 1; Length 31;  
Best Local Similarity 63.0%; Pred. No. 3.4e-07;  
Matches 17; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

OY 4 GTXXXXXKXKEEAVRLXXXXLKNG 30  
DB 4 GTFTSDLSKXKEEAVRLFTIEMLKNG 30

RESULT 5  
US-08-066-480-4  
Sequence 4, Application US/08066480  
Patent No. 5424286  
GENERAL INFORMATION:  
APPLICANT: Eng, John  
TITLE OF INVENTION: Pharmaceutical Compositions And Use of  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Allegritti & Witcoff, Ltd.  
STREET: 10 S. Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/066,480  
FILING DATE: 24-MAR-1993  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: McDonnell, John J  
REGISTRATION NUMBER: 26,949  
REFERENCE/DOCKET NUMBER: 93,084  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-715-1000  
TELEFAX: 312-715-1234  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 31 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..31  
OTHER INFORMATION: /label= y31-Extendin4  
US-08-066-480-4

Query Match 62.0%; Score 75; DB 1; Length 31;

Best Local Similarity 63.0%; Pred. No. 3.4e-07;  
Matches 17; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

OY 4 GTXXXXXKXKEEAVRLXXXXLKNG 30  
DB 4 GTFTSDLSKXKEEAVRLFTIEMLKNG 30

RESULT 6  
US-08-317-305-2  
Sequence 2, Application US/08317305  
Patent No. 5863744  
GENERAL INFORMATION:  
APPLICANT: Avraham, Shalom  
APPLICANT: Avraham, Hava  
TITLE OF INVENTION: NOVEL PROTEIN MARKER AND DNA ENCODING  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes  
STREET: Ten Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/317,305  
FILING DATE: 03-OCT-1994  
CLASSIFICATION: 436  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Ph.D., Kathleen A.  
REGISTRATION NUMBER: 34,380  
REFERENCE/DOCKET NUMBER: DH-001XX  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-2290  
TELEFAX: (617) 451-0313  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 589 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-317-305-2

Query Match 33.9%; Score 41; DB 2; Length 589;  
Best Local Similarity 45.0%; Pred. No. 12;  
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

OY 11 SKXKEEAVRLXXXXLKNG 30  
DB 262 SKIIVEAIRCKIKILQNDG 281

RESULT 7  
US-08-317-305-4  
Sequence 4, Application US/08317305  
Patent No. 5863744  
GENERAL INFORMATION:  
APPLICANT: Avraham, Shalom  
APPLICANT: Avraham, Hava  
TITLE OF INVENTION: NOVEL PROTEIN MARKER AND DNA ENCODING  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes

STREET: Ten Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/317,305  
FILING DATE: 03-OCT-1994  
CLASSIFICATION: 436  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Ph.D., Kathleen A.  
REGISTRATION NUMBER: 34,380  
REFERENCE/DOCKET NUMBER: DH-001XX  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-2290  
TELEFAX: (617) 451-0313  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 589 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-317-305-4

Query Match 33.9%; Score 41; DB 2; Length 589;  
Best Local Similarity 45.0%; Pred. No. 12;  
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 11 SKOXEEAVRLXXXLKNG 30  
II: |||:| |:  
DB 262 SKEIVEAIRCKIKLQNDG 281

RESULT 8  
US-08-862-508-2  
Sequence 2, Application US/08862508  
Patent No. 6066451  
GENERAL INFORMATION:  
APPLICANT: Avraham, Shalom  
APPLICANT: Groopman, Jerome E.  
TITLE OF INVENTION: NOVEL NEURAL CELL PROTEIN MARKER RR/B AND DNA  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lahive & Cockfield, LLP  
STREET: 28 State Street  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/862,508  
FILING DATE: 23-MAY-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/317,305  
FILING DATE: 03-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Silverl, Jean M.  
REGISTRATION NUMBER: 39,030  
REFERENCE/DOCKET NUMBER: NER-259DV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400

TELEFAX: (617) 742-4214  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 589 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-862-508-2

Query Match 33.9%; Score 41; DB 3; Length 589;  
Best Local Similarity 45.0%; Pred. No. 12;  
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 11 SKOXEEAVRLXXXLKNG 30  
II: |||:| |:  
DB 262 SKEIVEAIRCKIKLQNDG 281

RESULT 9  
US-08-862-508-4  
Sequence 4, Application US/08862508  
Patent No. 6066451  
GENERAL INFORMATION:  
APPLICANT: Avraham, Shalom  
APPLICANT: Groopman, Jerome E.  
TITLE OF INVENTION: NOVEL NEURAL CELL PROTEIN MARKER RR/B AND DNA  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lahive & Cockfield, LLP  
STREET: 28 State Street  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/862,508  
FILING DATE: 23-MAY-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/317,305  
FILING DATE: 03-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Silverl, Jean M.  
REGISTRATION NUMBER: 39,030  
REFERENCE/DOCKET NUMBER: NER-259DV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 742-4214  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 589 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-862-508-4

Query Match 33.9%; Score 41; DB 3; Length 589;  
Best Local Similarity 45.0%; Pred. No. 12;  
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 11 SKOXEEAVRLXXXLKNG 30  
II: |||:| |:  
DB 262 SKEIVEAIRCKIKLQNDG 281

RESULT 10

PCT-US95-12508-2  
; Sequence 2, Application PC/TUS9512508  
; GENERAL INFORMATION:  
; APPLICANT: Avraham, Shalom  
; APPLICANT: Avraham, Hava  
; APPLICANT: Groopman, Jerome E.  
; TITLE OF INVENTION: NOVEL PROTEIN MARKER AND DNA ENCODING SAME  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lahive & Cockfield  
; STREET: 60 State Street, Suite 510  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/12508  
; FILING DATE: 29-SEP-1995  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 08/317,305  
; FILING DATE: 03-OCT-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Silverl, Jean M.  
; REGISTRATION NUMBER: 39,030  
; REFERENCE/DOCKET NUMBER: NER-259PC  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 227-7400  
; TELEFAX: (617) 227-5941  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 589 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; PCT-US95-12508-2

Query Match 33.9%; Score 41; DB 4; Length 589;  
Best Local Similarity 45.0%; Pred. No. 12;  
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 11 SKXEEBAVRLXXXIKNG 30  
||: |||:| | :| |  
Db 262 SKEIVEAIRCKIKLONDG 281

RESULT 11  
PCT-US95-12508-4  
; Sequence 4, Application PC/TUS9512508  
; GENERAL INFORMATION:  
; APPLICANT: Avraham, Shalom  
; APPLICANT: Avraham, Hava  
; APPLICANT: Groopman, Jerome E.  
; TITLE OF INVENTION: NOVEL PROTEIN MARKER AND DNA ENCODING SAME  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lahive & Cockfield  
; STREET: 60 State Street, Suite 510  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/12508  
; FILING DATE: 29-SEP-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/317,305  
; FILING DATE: 03-OCT-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Silverl, Jean M.  
; REGISTRATION NUMBER: 39,030  
; REFERENCE/DOCKET NUMBER: NER-259PC  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 227-7400  
; TELEFAX: (617) 227-5941  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 589 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; PCT-US95-12508-4

Query Match 33.9%; Score 41; DB 4; Length 589;  
Best Local Similarity 45.0%; Pred. No. 12;  
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 11 SKXEEBAVRLXXXIKNG 30  
||: |||:| | :| |  
Db 262 SKEIVEAIRCKIKLONDG 281

RESULT 12  
US-08-062-024B-5  
; Sequence 5, Application US/08062024B  
; Patent No. 5708139  
; GENERAL INFORMATION:  
; APPLICANT: Alan Collmer and Sheng-Yang He  
; TITLE OF INVENTION: Pseudomonas syringae pv. syringae hrpZ Gene  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Yabwak & Associates  
; STREET: 25 Skytop Drive  
; CITY: Trumbull  
; STATE: Connecticut  
; COUNTRY: USA  
; ZIP: 06611  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Macintosh  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: Microsoft Word 4.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/062,024B  
; FILING DATE: May 17th 1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: George M. Yabwak  
; REGISTRATION NUMBER: 26,824  
; REFERENCE/DOCKET NUMBER: CRF D-1425  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (203)268-1951  
; TELEFAX: (203)268-1951  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 341 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-062-024B-5

Query Match 32.2%; Score 39; DB 1; Length 341;  
Best Local Similarity 32.0%; Pred. No. 15;



```

1  REGISTRATION NUMBER: 30,727
2  REFERENCE/DOCKET NUMBER: 14603/10050
3  TELECOMMUNICATION INFORMATION:
4  TELEPHONE: (716) 263-1304
5  TELEFAX: (716) 263-1600
6  INFORMATION FOR SEQ ID NO: 5:
7  SEQUENCE CHARACTERISTICS:
8  LENGTH: 341 amino acids
9  TYPE: amino acid
10 STRANDEDNESS:
11 TOPOLOGY: linear
12 MOLECULE TYPE: protein
13
14 US-08-819-559-5

```

Query Match	32.2%	Score 39	DB 2	Length 341
Best Local Similarity	32.0%	Pred. No. 15		
Matches	8	Conservative	4	Mismatches 13
				Indels 0
				Gaps 0
Qy	5	TXXXXXXKQEEANVRLXXXXLKG	29	
			:	:
Db	27	TTGCTSTKALQEVVVKLAELMNG	51	

Search completed: February 13, 2001, 16:07:46  
Job time: 101 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2001, 16:08:13 ; Search time 21.65 Seconds

(without alignments)  
125.452 Million cell updates/sec

Title: US-08-908-867-39

Perfect score: 121  
Sequence: 1 XXXGTXXXXSKOXEEPAVRLXXXLKNKGSGAGXXXXX 40

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : PIR\_66:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	91	75.2	39	1 HMGH32	extendin-3 - Mexica
2	91	75.2	39	1 HMGH32	extendin-4 - Gila m
3	42	34.7	546	2 G6A803	phosphoglucumtase
4	41	33.9	141	2 T39165	conserved hypotet
5	41	33.9	157	2 G75266	hypothetical prote
6	41	33.9	402	2 A75054	molybdenum cofacto
7	41	33.9	2044	2 T13704	still life protein
8	41	33.9	2064	2 T13707	still life protein
9	40	33.1	127	2 C69774	transcription regu
10	40	33.1	609	2 T45637	beta-D-glucan exch
11	40	33.1	772	2 T06154	hypothetical prote
12	39	32.2	208	2 D71137	probable transcrip
13	39	32.2	341	2 A40706	extracellular hype
14	39	32.2	688	2 E71845	polynucleotide pho
15	39	32.2	688	2 E64671	polynucleotide pho
16	38.5	31.8	653	2 T02080	probable carbonate
17	38.5	31.8	1702	2 T14050	protein kinase (EC
18	38	31.4	300	2 E71023	transcription init
19	38	31.4	300	2 E75110	transcription init
20	38	31.4	357	2 TCA703	basic helix-loop-h
21	38	31.4	357	2 T49338	neurogenic differe
22	38	31.4	381	2 A57059	beta-cell E-box tr
23	38	31.4	636	2 T45640	beta-D-glucan exch
24	38	31.4	636	2 T45640	beta-D-glucan exch
25	38	31.4	726	2 T20183	hazooka gene prote
26	38	31.4	1464	2 T13716	hazooka gene prote
27	37	30.6	430	2 S50604	AS12 protein - yea
28	37	30.6	488	2 S40706	hypothetical prote
29	37	30.6	624	2 T04414	probable glucan 1,

30	37	30.6	628	2 T51283	glucan 1,3-beta-gl
31	37	30.6	850	2 T13352	stn-A protein - Ir
32	37	30.6	938	2 T05533	hypothetical prote
33	37	30.6	1237	2 A31334	phosphorylase kina
34	36.5	30.2	1314	2 T09481	matting type silenc
35	36	29.8	85	2 H83399	hypothetical prote
36	36	29.8	115	2 S57269	hypothetical prote
37	36	29.8	153	2 T29164	hypothetical prote
38	36	29.8	160	2 A69849	hypothetical prote
39	36	29.8	238	2 D71189	hypothetical prote
40	36	29.8	268	1 J00961	myb-related protei
41	36	29.8	284	2 J06198	alpha-tropomyosin
42	36	29.8	303	1 S57582	methanol dehydroge
43	36	29.8	318	2 A43746	nisin resistance p
44	36	29.8	324	2 C42514	H3L protein - vacc
45	36	29.8	324	2 T37369	IMV membrane assoc

#### ALIGNMENTS

RESULT 1  
HMGH32  
extendin-3 - Mexican beaded lizard  
C:Species: Heloderma horridum (Mexican beaded lizard)  
C>Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 21-Nov-1997  
C:Accession: A23674  
R:Eng, J.; Andrews, P.C.; Kleinman, W.A.; Singh, L.; Raufman, J.P.  
J: Biol. Chem. 265, 20259-20262, 1990  
A:Title: Purification and structure of extendin-3, a new pancreatic secretagogue isola  
A:Reference number: A23674; MUID:91056067  
A:Accession: A23674  
A:Residues: 1-39 <ENG>  
A:Molecule type: protein  
C:Comment: Extendins are venom components that are thought to bind to receptors for va  
g in secretion of amylin.  
C:Superfamily: glucagon  
C:Keywords: amidated carboxyl end; duplication; secretagogue; venom  
F:39/Modified site: amidated carboxyl end (Ser) #status experimental

Query Match 75.2%; Score 91; DB 1; Length 39;  
Best Local Similarity 65.6%; Pred. No. 5.6e-10;  
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
QY 4 GTXXXXXSKOXEEPAVRLXXXLKNKGSSGA 35  
Db 4 GTFTSDLSKOMEPAVRLFIEMLKNKGSSGA 35

RESULT 2  
HMGH32  
extendin-4 - Gila monster  
C:Species: Heloderma suspectum (Gila monster)  
C>Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 21-Nov-1997  
C:Accession: A42486  
R:Eng, J.; Kleinman, W.A.; Singh, L.; Singh, G.; Raufman, J.P.  
J: Biol. Chem. 267, 7402-7405, 1992  
A:Title: Isolation and characterization of extendin-4, an extendin-3 analogue, from Hel  
A:Reference number: A42486; MUID:92218391  
A:Accession: A42486  
A:Molecule type: protein  
A:Residues: 1-39 <ENG>  
C:Comment: Extendin-4 does not stimulate amylase secretion by pancreatic acinar cells.  
C:Superfamily: glucagon  
C:Keywords: amidated carboxyl end; duplication; venom  
F:39/Modified site: amidated carboxyl end (Ser) #status experimental

Query Match 75.2%; Score 91; DB 1; Length 39;  
Best Local Similarity 65.6%; Pred. No. 5.6e-10;  
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY            4 GTXXXXXSKQEEEAARLXXXXLKNGCASSGA 35  
             ||     ||| |||||     ||||| ||||  
Db            4 GTFTSDLSKQEEEAARLFIEWLKNGCPSSGA 35

### RESULT 3

phosphoglucutransferase (EC 5.4.2.2) - *Escherichia coli*  
C:Species: *Escherichia coli*  
C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 17-Mar-2000  
C:Accession: G64803; I55076  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co-  
A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of *Escherichia coli* K-12.  
A:Reference number: A64720; MUID:9742617  
A:Accession: G64803  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-546 <BLAT>  
A:Cross-references: GB:AF000172; GS:U00096; NID:g17869896; PIDN:AAC73782.1; PID:g1786904.1  
R:Experimental source: strain K-12, substrain MG1655  
R:Lu, M.; Kleckner, N  
J. Bacteriol. 176, 5847-5851, 1994  
A:Title: Molecular cloning and characterization of the *pgm* gene encoding phosphoglucutrans-  
A:Reference number: I55076; MUID:94364967  
A:Accession: I55076  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-21, 'M', 23-546 <RES>  
A:Cross-references: EMBL:U008369; NID:g473887; PIDN:AAA57067.1; PID:g473888  
C:Genetics:  
A:Gene: *pgm*  
A:Function:  
A:Description: conversion of D-glucose 1-phosphate into D-glucose 6-phosphate; participle  
C:Superfamily: phosphoglucutransferase  
C:Keywords: intramolecular transferase; isomerase; phosphoprotein  
E146/Active site: Ser (phosphoserine intermediate) #status predicted

Query Match	34.7%	Score 42	DB 2	length 546
Best Local Similarity	52.9%	Pred. No. 8.9'		
Matches	9	Conservative	2	Mismatches 6; Indels 0; Gaps 0.
QY	12	KXEBEAVRLXXXXLKN	28	
DB	539	KXIEKEAVEIVLKN	545	

## RESULT 4

conserved hypothetical protein SPAC8E11.11 - fls180n yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: J39165  
R:McLean, J.; Harris, D.; Wood, V.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, February 1998  
A:Reference number: Z1831  
A:Accession: J39165  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-141 <MC1>  
A:Cross-references: EMBL:AL021817; PIN:CA940196.1; GSPDB:GND0066; SPDB:SPAC8E11.11  
A:Experimental source: strain 972h-; cosmid c8E11  
C:Genetics:  
A:Gene: SPDB:SPAC8E11.11  
A:Map position: 1

Query Match	33.9%	Score 41;	DB 2;	Length 141;
Best Local Similarity	34.8%	Pred. No. 3.2;		
Matches	8;	Conservative	5;	Mismatches 10;
				Indels 0;
				Gaps 0;

```
QY 12 KQEEEA VRLXXXXLKN GXS5G 34
    | : | : | : | | | : |
DB 64 KETEVA IEVTKWILSNGGVWNG 86
```

3  
RESULT  
G75266

hypothetical protein DR2500 - Deinococcus radiodurans (strain R1)  
C:Species: deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
C:Accession: G75266  
R:White, O.; Eissen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M.: Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uitterback, T.; Zalewski, C.;  
S.: Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896  
A:Accession: G75266  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-157 <WHI>  
A:Cross-references: GB:AE002079; GB:AE000513; NID:6460315; PIDN:AF12045.1; PID:9646  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DR2500  
A:Map position: 1  
C:Superfamily: Deinococcus radiodurans hypothetical protein DR2500

Query Match	33.98;	Score 41;	DB 2;	Length 157;
Best Local Similarity	42.18;	Pred. No. 3.6;		
Matches	8;	Conservative	4;	Mismatches 7;
				Indels 0;
				Gaps 0

```
QY      16  EEA VRLXXXXLKNGGXSSG  34
          ::||:  |||  |  ||
Db      74  DDAVQVFYRALKNAGLDSG  92
```

## RESULT

molybdenum cofactor biosynthesis protein (moea-1) PAB1436 - Pyrococcus abyssi (strain  
 C:Species: Pyrococcus abyssi  
 C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
 C:Accession: A75054  
 R:Anonymous, Genoscope  
 submitted to the EMBL Data Library, July 1999  
 A:Description: Pyrococcus abyssi genome sequence: Insights into archaeal chromosome s  
 A:Reference number: A75001  
 A:Accession: A75054  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-402 <KAW>  
 A:Cross-references: GB:AB248287; GB:AL096836; NID:95459657; PIDN:CAB50326.1; PID:9545  
 A:Experimental source: strain Orsay  
 C:Genetics:  
 A:Gene: PAB1436  
 C:Superfamily: molybdenum cofactor biosynthesis protein moea-2

Query Match	33.9%	Score 41;	DB 2;	Length 402;
Best Local Similarity	39.1%;	Pred. No. 9.8;		
Matches	9;	Conservative	4;	Mismatches 10;
			Indels	0;
			Gaps	0

```
QY      12 KQXEEAVRLXXXXLKNGGXSSG 34
        | : | | | : : | | |
Db      237 KELIEGVRVADIVVISGASGG 259
```

## RESULT

13/04  
still life protein type 2 - fruit fly (*Drosophila melanogaster*)  
C\_Species: *Drosophila melanogaster*  
C\_Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 21-Jul-2000

C:Accession: T13704  
 R:Some, M.; Hoshino, M.; Suzuki, E.; Kuroda, S.; Kaibuchi, K.; Nakagoshi, H.; Saigo, K.; Science 275, 543-547, 1997  
 A:Title: Still life, a protein in synaptic terminals of Drosophila homologous to GDP-GTf  
 A:Reference number: Z17701; MUID:97153054  
 A:Accession: T13704  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-2044 <SON>  
 A:Cross-references: EMBL:D86546; NID:g1813375; PIDN:BAI13108.1; PID:g1813376

Query Match 33.9%; Score 41; DB 2; Length 2044;  
 Best Local Similarity 41.7%; Pred. No. 55;  
 Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 12 KQEEAVRLXXXXLKNKGXSGA 35  
 DB 1761 RQIRRESVRNMSIPMKNFSGSSG 1784

RESULT 8  
 T13707  
 still life protein type 1 - fruit fly (Drosophila melanogaster)  
 C:Species: Drosophila melanogaster  
 C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 21-Jul-2000  
 C:Accession: T13707  
 R:Some, M.; Hoshino, M.; Suzuki, E.; Kuroda, S.; Kaibuchi, K.; Nakagoshi, H.; Saigo, K.; Science 275, 543-547, 1997  
 A:Title: Still life, a protein in synaptic terminals of Drosophila homologous to GDP-GTf  
 A:Reference number: Z17701; MUID:97153054  
 A:Accession: T13707  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-2064 <SON>  
 A:Cross-references: EMBL:D86547; NID:g1813377; PIDN:BAI13109.1; PID:g1813378

Query Match 33.9%; Score 41; DB 2; Length 2064;  
 Best Local Similarity 41.7%; Pred. No. 55;  
 Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 12 KQEEAVRLXXXXLKNKGXSGA 35  
 DB 1781 RQIRRESVRNMSIPMKNFSGSSG 1804

RESULT 9  
 C69774  
 transcription regulator phage-related homolog ydcN - Bacillus subtilis  
 C:Species: Bacillus subtilis  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000  
 C:Accession: C69774  
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Allout, G.; Azevedo, V.; Berter, C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd A.; Erlich, S.D.; Emmerison, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E. Nature 390, 249-256, 1997  
 A:Authors: Foulter, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Gallier, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinols, A.; Laubert, P.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle Rieger, M.; Rivolta, C.; Rocca, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot akouchi, M.; Tanakoshi, A.; Tanaka, T.; Terpeira, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
 A:Reference number: A69580; MUID:98044033  
 A:Accession: C69774  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-127 <KUN>

A:Cross-references: GB:299106; GB:AL009126; NID:g2632653; PIDN:CAB12289.1; PID:g26327  
 A:Experimental source: strain 168  
 C:Genetics:  
 A:Gene: ydcN  
 C:Superfamily: probable transcription repressor yowR

Query Match 33.1%; Score 40; DB 2; Length 127;  
 Best Local Similarity 47.1%; Pred. No. 4.4;  
 Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 13 QXEEAVRLXXXXLKNKG 29  
 DB 100 EFDEETARLVKKALKNKG 116

RESULT 10  
 T45637  
 beta-D-glucan exohydrolase-like protein - Arabidopsis thaliana  
 N:Alternate names: protein F13112.60  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 21-Jul-2000  
 C:Accession: T45637  
 R:Choisme, N.; Robert, C.; Brothier, P.; Wincker, P.; Catolico, L.; Artiguenave, F.; submitted to the Protein Sequence Database, November 1999  
 A:Reference number: Z23010  
 A:Accession: T45637  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-609 <CHO>  
 A:Cross-references: EMBL:AL133292  
 A:Experimental source: cultivar Columbia; BAC clone F13112  
 C:Genetics:  
 A:Map position: 3  
 A:Introns: 57/1; 125/2; 155/3; 204/2; 285/3; 320/3; 381/3; 449/1  
 A:Note: F13112.60  
 C:Superfamily: beta-glucosidase

Query Match 33.1%; Score 40; DB 2; Length 609;  
 Best Local Similarity 38.5%; Pred. No. 23;  
 Matches 10; Conservative 2; Mismatches 14; Indels 0; Gaps 0;

QY 4 GTXXXXSKQEEAVRLXXXXLKNKG 29  
 DB 376 GTVCKEHEVAREAVRKSVILKNKG 401

RESULT 11  
 T06154  
 hypothetical protein F24J7.162 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 22-Oct-1999  
 C:Accession: T06154  
 R:Beyan, M.; Vitale, D.; Liqouri, R.; Argitrou, A.; De Simone, V.; Bancroft, I.; Mewe submitted to the Protein Sequence Database, April 1999  
 A:Reference number: Z15493  
 A:Accession: T06154  
 A:Molecule type: DNA  
 A:Residues: 1-772 <BEV>  
 A:Cross-references: EMBL:AL021768; GSPDB:GN00062; ATSP:F24J7.162  
 A:Experimental source: cultivar Columbia; BAC clone F24J7  
 C:Genetics:  
 A:Gene: ATSP:F24J7.162  
 A:Map position: 4  
 A:Introns: 4/2; 42/3; 273/2; 303/2; 342/3; 346/1; 463/3; 485/2; 536/3; 548/3; 576/3;  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-127 <KUN>

Query Match 33.1%; Score 40; DB 2; Length 772;  
 Best Local Similarity 37.5%; Pred. No. 30;  
 Matches 9; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 11 SKQEEAVRLXXXXLKNKGXSSG 34

Db 60 SHKKEEAKKSSSEGLKDGNAKGG 83

# RESULT 12

D71137 probable transcription initiation factor IIB - *Pyrococcus horikoshii*

C:Species: *Pyrococcus horikoshii*

C:Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 21-Jul-2000

C:Accession: D71137

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin

M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushiida, N.; Oguchi

DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic

A:Reference number: A71000; MUID:98344137

A:Accession: D71137

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-208 <KAM>

A:Cross-references: GB:AP000003; NID:g3236130; PIDN:BA29958.1; PID:g3257275

A:Experimental source: strain OT3

A:Note: this accession replaces an interim accession for a sequence replaced by GenBank

C:Genetics:

A:Gene: PH0864

C:Superfamily: transcription initiation factor IIB; transcription initiation factor IIB

C:Keywords: transcription initiation

Query Match 32.2%; Score 39; DB 2; Length 208;  
Best Local Similarity 44.4%; Pred. No. 11;

Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 12 SKXEEAVRLXXXXLKNKG 29

Db 38 KHVEREAVRLVRLKNSG 55

# RESULT 13

A40706

extracellular hypersensitive necrosis response elicitor, 34.7K - *Pseudomonas syringae*

C:Species: *Pseudomonas syringae*

C:Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 08-Oct-1999

C:Accession: A40706

R:He, S.Y.; Huang, H.C.; Collier, A.

Cell 73, 1255-1266, 1993

A:Title: *Pseudomonas syringae* pv. *syringae* harpinps: a protein that is secreted via the

A:Reference number: A40706; MUID:933133957

A:Accession: A40706

A>Status: preliminary

A:Molecule type: DNA; protein

A:Residues: 1-341 <HEIL>

A:Cross-references: GB:L14775; NID:g349793; PIDN:AA25839.1; PID:g349794

A:Experimental source: *syringae* 61

A:Note: sequence extracted from NCBI backbone (NCBIN:135085, NCBIP:135086)

Query Match 32.2%; Score 39; DB 2; Length 341;  
Best Local Similarity 32.0%; Pred. No. 19;

Matches 8; Conservative 4; Mismatches 13; Indels 0; Gaps 0;

QY 5 TXXXXXSKXEEAVRLXXXXLKNKG 29

Db 27 TTGTSSKALQEVVVKLAELMRNG 51

# RESULT 14

E71845

polynucleotide nucleotidyltransferase - *Helicobacter pylori* (strain J99)

C:Species: *Helicobacter pylori*

A:Variety: strain J99

C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 24-Sep-1999

C:Accession: E71845

R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Dolg, P.C.; Smith, D.R.

; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Voyts, G.F

Nature 397, 176-180, 1999

A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric p

A:Reference number: A71800; MUID:99120557

A:Accession: E71845

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-688 <ARN>

A:Cross-references: GB:AE001541; GB:AE001439; NID:g4155724; PIDN:AD06718.1; PID:g415

A:Experimental source: strain J99

C:Genetics:

A:Gene: pnp

C:Superfamily: polynucleotide nucleotidyltransferase alpha chain

Query Match 32.2%; Score 39; DB 2; Length 688;  
Best Local Similarity 36.0%; Pred. No. 40;

Matches 9; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 11 SKXEEAVRLXXXXLKNKGXSSGA 35

Db 415 SIKKQVIRLVSEILNNGSSSMA 439

# RESULT 15

E64671 polynucleotide phosphotriase - *Helicobacter pylori* (strain 26695)

C:Species: *Helicobacter pylori*

C:Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 24-Sep-1999

C:Accession: E64671

R:Tom, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R

Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKe

son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Wathey,

Nature 388, 539-547, 1997

A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karp, P.D.; Smith, H.O.; Fraser,

A:Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.

A:Reference number: A64520; MUID:97394467

A:Accession: E64671

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-688 <TON>

A:Cross-references: GB:AE000627; GB:AE000511; NID:g231373; PIDN:AD08258.1; PID:g231

C:Superfamily: polynucleotide nucleotidyltransferase alpha chain

Query Match 32.2%; Score 39; DB 2; Length 688;  
Best Local Similarity 36.0%; Pred. No. 40;

Matches 9; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 11 SKXEEAVRLXXXXLKNKGXSSGA 35

Db 415 SIKKQVIRLVSEILNNGSSSMA 439

Search completed: February 13, 2001, 16:08:14  
Job time: 129 sec



**THIS PAGE BLANK (USPTO)**

No.	Score	Match	Length	DB	ID	Description
1	91	75.2	39	1	EXE3_HELHO	P20394 heloderma h
2	91	75.2	87	1	EXE4_HELST	P26349 heloderma s
3	42	34.7	546	1	PGWJ_ECOLI	P26938 escherichia l
4	41	33.9	516	1	PI10_HOMAN	O14662 homo sapien
5	41	33.9	2044	1	PI12_DROME	P16160 drosophila
6	41	33.9	2064	1	SIPL_DROME	P16161 drosophila
7	39	32.2	341	1	HRPZ_PSESY	P35674 pseudomonaa
8	38	31.4	300	1	TEPZ_PYRHO	059151 pyrococcus
9	38	31.4	355	1	NDFI_MESAU	060401 mesocricetu
10	38	31.4	357	1	NDFI_MOUSE	060867 mus musculu
11	38	31.4	357	1	NDFI_RAT	064269 rattus norv
12	38	31.4	419	1	DNLI_ASEFM	P26813 artican swi
13	38	31.4	589	1	LU1A_LYCNP	O04973 lycopersicc
14	37	30.6	430	1	ASRZ_YEAST	P39945 saccharomyc
15	37	30.6	488	1	YKRT_CABEL	P44312 caenorhabdi
16	37	30.6	1237	1	KPBI_RABIT	P18668 cryptotolagu
17	36	29.8	318	1	NSR_LACLA	P23668 lactococcus
18	36	29.8	324	1	GLXK_RHIME	067369 rhizobium n
19	36	29.8	324	1	VP35_VACCC	P20497 vaccinia vi
20	36	29.8	335	1	VP35_VARY	P33059 variola vir
21	36	29.8	373	1	B1OF_AOUAE	O66875 aquilex aec
22	36	29.8	401	1	CXPX_BRAUA	059203 bradrythzoc
23	36	29.8	413	1	FIL_TOBAC	O40504 nicotiana t
24	36	29.8	633	1	SHRH_RHIME	P15715 rhizobium n
25	36	29.8	845	1	SCPL_MESAU	O60563 mesocricetu
26	36	29.8	3068	1	POLG_MESAU	O01500 p genome p
27	36	29.8	4687	1	PLRC_RAT	P30427 rattus norv
28	35	29.3	111	1	HMGZ_DROME	O06943 drosophila
29	35	28.9	273	1	TEPZ_PYRMO	P29045 streptococc
30	35	28.9	261	1	EPGZ_STRMU	P55005 streptococc
31	35	28.9	300	1	TEPZ_PYRPU	O51731 pyrococcus
32	35	28.9	356	1	NDFI_HOMAN	O43562 homo sapien
33	35	28.9	416	1	FLZ2_TOBAC	P40505 nicotiana t

45	35	28.9	460	1	VP41_BPAS	Q9187	bacterioph
34	35	28.9	467	1	HEU1_PACB	P39778	bacillus su
35	35	28.9	501	1	FLA4_MOUSE	P46620	aquifer pyr
36	35	28.9	509	1	GARB_MOUSE	O33107	mycobacteri
37	35	28.9	509	1	GARB_MOUSE	O33107	mycobacteri
38	35	28.9	563	1	ID5_MOUSE	O08880	mus musculu
39	35	28.9	639	1	CA1C_RABIT	O28902	oryctolagus
40	35	28.9	748	1	HEPA_HSVBC	P52374	bovine herp
41	35	28.9	759	1	NAH6_ONCAY	O01345	oncorhynchu
42	35	28.9	858	1	H105_CRICK	O60446	cricetulus
43	35	28.9	938	1	MM1_HUMAN	O05566	homo sapien
44	35	28.9	938	1	NM21_MOUSE	P35458	mus musculu
45	35	28.9	938	1	NM21_NAT	P35459	rattus norv

## ALIGNMENTS

RESULT	1			
EXE3_HELHO				
ID	EXE3_HELHO	STANDARD;	PRT;	39 AA.
AC	P20354;			
DT	01-FEB-1991 (rel. 17, Created)			
DT	01-FEB-1991 (rel. 17, Last sequence update)			
DT	01-MAY-1992 (rel. 22, Last annotation update)			
DE	EXENDIN-3.			
OS	Holoderma horridum horridum (Mexican beaded lizard).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Lepidosauria; Squamata; Scleroglossae; Anguimorpha; Helodermatidae			
NC	Holoderma.			
NC	111			

RC TISSUE=VENOM;  
RX MEDLINE=9105067; PubMed=1700785;  
RA Edg J., Andrew P.C., Kleinman W.A., Singh L., Raufman J.-P.,  
RT "Purification and structure of exendin-3, a new pancreatic  
RT secretagogue isolated from *Holodermis horridum* venom."  
J Biol Chem. 265:20255-20263(1990)  
CC -1- FUNCTION: HAS A VIP/SECRETIN-LIKE BIOLOGICAL ACTIVITY. INTERACTS  
CC WITH THE EXENDIN RECEPTOR.  
CC -1- SIMILARITY: BELONGS TO THE GLUCAGON FAMILY.

DR PIR; A236/4; HMG332.  
DR HSP; P01274; IGCN.  
DR INTERPRO; IPR000532; -  
DR PFAM; PF00123; hormonez; 1.  
DR PROSITE; PS00260; GLUCAGON; 1.  
DR Glucagon family; Venom; Amidation.  
FT MOD\_RES 39  
FT 39  
FT SEQUENCE 39 AA; 4204 MW; A44251DBA451D1B9 CRC64;

Query Match	75.28;	Score 91;	DB 1;	Length 39;
Best Local Similarity	65.68;	Pred. No. 4.7e-10;		
Matches 21; Conservative	0;	Mismatches 11;	Indels 0;	Gaps 0

QY 4 GTXXXXXSKQEEAVRLXXXXLKNGGXSSGA 35  
|| ||| ||| ||| ||| ||| ||| |||  
Db 4 GFTSDLSKQMEEAVALFIWLKNGCPSSGA 35

RESULT	2		
EXEA_HELSTU			
ID	EXEA_HELSTU	STANDARD:	PRT; 87 AA.
AC	P26349;		
DT	01-MAY-1992 (rel. 22. Created)		
DT	15-JUL-1998 (rel. 36, Last sequence update)		
DT	30-MAY-2000 (rel. 39, Last annotation update)		
DE	EXENDIN-4 PRECURSOR.		
OS	Heloderma suspectum (Gila monster).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Lepidosauria; Squamata; Scleroglossae; Anguimorpha; Helodermatidae		
OC	Heloderma.		
RP	[1]		
RP	SEQUENCE FROM N.A.		

```

RX MEDLINE-97172477; PubMed-9020121;
RA Chen Y.E., Drucker D.J.;
RT "Tissue-specific expression of unique mRNAs that encode proglucagon-
RT derived peptides or extendin 4 in the lizard.";
RL J. Biol. Chem. 272:4108-4115(1997).
RN [2]
RP SEQUENCE OF 48-86.
RC TISSUE-VENOM;
RX MEDLINE-92218391; PubMed-1313797;
RA Eng J., Kleinman W.A., Singh L., Singh G., Raufman J.-P.;
RT "Isolation and characterization of extendin-4, an extendin-3 analogue,
RT from Heloderma suspectum venom. Further evidence for an extendin
RT receptor on dispersed acini from guinea pig pancreas.";
RL J. Biol. Chem. 267:7402-7405(1992).
CC -1- FUNCTION: HAS A VIP/SECRETIN-LIKE BIOLOGICAL ACTIVITY. INTERACTS
CC WITH THE EXTENDIN RECEPTOR.
CC -1- SIMILARITY: BELONGS TO THE GLUCAGON FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U77613; AAB51130.1; -
DR PIR: A42486; HMCHAG;
DR INTERPRO: IPR000532; -
DR PFM: PF00123; hormone; 1.
DR PROSITE: PS00260; GLUCAGON; 1.
KW Glucagon family; Venom; Amidation; Signal.
FT SIGNAL 1 23
FT PEPTIDE 48 86
FT MOD_RES 86 86
FT SEQUENCE 87 AA; 9479 MW; 656BA6E3D87454A2 CRC64;
SQ
Query Match 75.2%; Score 91; DB 1; Length 87;
Best Local Similarity 65.6%; Pred. No. 11e-09;
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 4 GTXXXXXKXQXEBAYRLXXXXLNKXSSGA 35
DB 51 GTFTSDLSKQWEEAVRLFIEWLKNGGSSGA 82

```

```

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RT Science 277:1453-1474(1997).
RL [3]
RN SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE-97061202; PubMed-8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitegawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiiuchi T.;
RT "A 718-Kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN [4]
RP SEQUENCE OF 1-20 FROM N.A.
RC STRAIN-K12;
RX MEDLINE-94236685; PubMed-8011018;
RA Lu M., Campbell J.L., Boye E., Kleckner N.;
RT "Seq. a negative modulator of replication initiation in E. coli.";
RL Cell 77:413-426(1994).
RN [5]
RP CHARACTERIZATION.
RA Josh J.G., Handler P.;
RT "Phosphoglucosyltransferase. II. Purification and properties of
RT phosphoglucosyltransferase from Escherichia coli.";
RL J. Biol. Chem. 239:2741-2751(1964).
CC -1- FUNCTION: THIS ENZYME PARTICIPATES IN BOTH THE BREAKDOWN AND
CC SYNTHESIS OF GLUCOSE.
CC -1- CATALYTIC ACTIVITY: ALPHA-D-GLUCOSE 1-PHOSPHATE = ALPHA-D-GLUCOSE
CC 6-PHOSPHATE.
CC -1- SIMILARITY: TO OTHER PHOSPHOGLUCOSYLTRANSFERASES AND PHOSPHOMANNOMUTASES.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U08369; AAA57067.1; -
DR EMBL: AE000172; AAC73782.1; -
DR EMBL: D90707; BAA35337.1; -
DR EMBL: D90708; BAA35345.1; -
DR EMBL: U07651; -; NOT_ANNOTATED_CDS.
DR ECOGENE: EG12144; PGM.
DR INTERPRO: IPR001485; -
DR PFM: PF00408; PGM_PMM; 1.
DR PROSITE: PS00710; PGM_PMM; 1.
KW isomerase; Phosphorylation.
FT ACT_SITE 146 146
FT FT
SQ SEQUENCE 546 AA; 58361 MW; 666B6B9C2F2ECD59 CRC64;
QY 12 KXQEEAVRLXXXXLNK 28
DB 529 KXQEEAVRLXXXXLNK 545

```

```

RESULT 4
ID P110_HUMAN STANDARD; PRT; 516 AA.
AC 014682;
DT 30-MAY-2000 (Rel. 39, Created)
Query Match 34.7%; Score 42; DB 1; Length 546;
Best Local Similarity 52.9%; Pred. No. 4.8;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

```



DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE P53-INDUCED PROTEIN 10.  
 GN PIG10.  
 GN Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-COLON CANCER;  
 RX MEDLINE:97449378; PubMed:9305847;  
 RA Poljak K., Xia Y., Zweiler J.L., Kinzler K.W., Vogelstein B.;  
 RT Nature 389:300-306(1997).  
 RL "A model for p53-induced apoptosis."  
 CC -1- SIMILARITY: BELONGS TO THE KELCH/MIPP FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF010314; AAC39532.1; -  
 DR INTERPRO: IPR000169; -  
 DR INTERPRO: IPR000210; -  
 DR INTERPRO: IPR001798; -  
 DR PFAM: PF00651; BTB; 1.  
 DR PFAM: PF01344; Kelch; 3.  
 DR PROSITE: PS50097; BTB; 1.  
 DR KW Repeat.  
 FT DOMAIN 30 147 BTB.  
 FT DOMAIN 299 447 3 APPROXIMATE KELCH TANDEM REPEATS.  
 FT REPEAT 299 343 1.  
 FT REPEAT 344 391 2.  
 FT REPEAT 392 447 3.  
 FT SEQUENCE 516 AA; 58314 MW; DFC3D777B3D3E9B2 CRC64;  
 SO  
 Query Match 33.9%; Score 41; DB 1; Length 516;  
 Best Local Similarity 45.0%; Pred. No. 6.9;  
 Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;  
 QY 11 SKXEEAVRLXXXXLKNKG 30  
 DB 265 SKXEEAVRLXXXXLKNKG 284  
 ID SIF2\_DROME STANDARD; PRT; 2044 AA.  
 AC P91620;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE STILL LIFE PROTEIN TYPE 2 (SIF TYPE 2).  
 GN SIF.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RX MEDLINE:97153054; PubMed:8999801;  
 RA Sone M., Hoshino M., Suzuki E., Kuroda S., Kaibuchi K., Nakagoshi H.,  
 RA Saigo K., Nabeshima Y.-I., Hama C.;  
 RT "Still life", a protein in synaptic terminals of Drosophila homologous  
 RT to GDP-GTP exchangers."  
 RL Science 275:543-547(1997).  
 RN [2]

RP ERRATUM.  
 RA Sone M., Hoshino M., Suzuki E., Kuroda S., Kaibuchi K., Nakagoshi H.,  
 RA Saigo K., Nabeshima Y.-I., Hama C.;  
 RL Science 275:1405-1405(1997).  
 CC -1- FUNCTION: REGULATES SYNAPTIC DIFFERENTIATION THROUGH THE  
 CC ORGANIZATION OF ACTIN CYTOSKELETON POSSIBLY BY ACTIVATING RHO-LIKE  
 CC GTPASES. IS LIKELY A FACTOR IN THE CASCADE OF RAC1 OR CDC42 IN THE  
 CC NEURONS.  
 CC -1- SUBCELLULAR LOCATION: LOCALIZES TO THE SUBMEMBRANOUS REGION OF  
 CC SYNAPTIC TERMINALS.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: SIF TYPE 1 (P91621) AND SIF TYPE  
 CC 2 (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- DEVELOPMENTAL STAGE: AT STAGE 14, EXPRESSION OCCURS IN EACH  
 CC SEGMENT OF THE CENTRAL NERVOUS SYSTEM. AT STAGE 17, EXPRESSION  
 CC BECOMES RESTRICTED TO THE SYNAPTIC REGIONS OF THE BRAIN AND  
 CC VENTRAL NERVE CORD, WHERE SYNAPSES UNDERGO MATURATION.  
 CC -1- SIMILARITY: CONTAINS 1 DBL-HOMOLOG DOMAIN (DH).  
 CC -1- SIMILARITY: CONTAINS 2 PH DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D86546; BAA13108.1; -  
 DR FLYBASE; FBgn0019652; SIF.  
 DR INTERPRO: IPR000219; -  
 DR INTERPRO: IPR001311; -  
 DR INTERPRO: IPR001478; -  
 DR INTERPRO: IPR001849; -  
 DR PFAM: PF00595; PDZ; 1.  
 DR PFAM: PF00169; PH; 2.  
 DR PFAM: PF00621; Rhogef; 1.  
 DR PROSITE: PS00741; GDS\_CDC24; 1.  
 DR PROSITE: PS50003; PH\_DOMAIN; 1.  
 DR KW Guanine-nucleotide releasing factor; Developmental protein; Synapse;  
 KW Repeat; Alternative splicing.  
 FT DOMAIN 62 249 4 X 25 AA APPROXIMATE REPEAT.  
 FT REPEAT 62 86 1.  
 FT REPEAT 94 118 2.  
 FT REPEAT 154 178 3.  
 FT REPEAT 225 249 4.  
 FT DOMAIN 819 937 PH.  
 FT DOMAIN 1184 1273 PDZ.  
 FT DOMAIN 1410 1673 DH.  
 FT DOMAIN 1674 1767 PH.  
 FT DOMAIN 467 470 POLY-PRO.  
 FT DOMAIN 646 649 POLY-ARG.  
 FT DOMAIN 1295 1298 POLY-PRO.  
 FT DOMAIN 1898 1909 POLY-GLN.  
 FT DOMAIN 1929 1933 POLY-PRO.  
 SO SEQUENCE 2044 AA; 228324 MW; 75D7CF21F49654B6 CRC64;  
 QY 12 KOXEEAVRLXXXXLKNKGSSGA 35  
 DB 1761 KOXEEAVRLXXXXLKNKGSSGA 1784  
 ID SIF1\_DROME STANDARD; PRT; 2064 AA.  
 AC P91621;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 RN [2]



[1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-OT3;  
 RC MEDLINE-98344137; PubMed-9679194;  
 RA Kawarabayashi Y., Sawada M., Horikawa H., Hatakeyama Y., Hino Y.,  
 RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,  
 RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohnuki Y.,  
 RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A.,  
 RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,  
 RA Maschl Y., Shizuya H., Kikuchi H.;  
 RT "Complete sequence and gene organization of the genome of a hyper-  
 thermophilic archaebacterium, Pyrococcus horikoshii OT3.";  
 RL DNA Res. 5:55-76(1998).  
 CC -1- FUNCTION: STABILIZES TBP BINDING TO AN ARCHAEAL BOX-A PROMOTER.  
 CC ALSO RESPONSIBLE FOR RECRUITING RNA POLYMERASE II TO THE PRE-  
 CC INITIATION COMPLEX (DNA-TBP-TFIIB) (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE TFIIB FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: AP000006; BAA30589.1; -  
 DR INTERPRO: IPR000812; -  
 DR PFAM: PF00382; transcript\_fac2; 2.  
 DR PRINTS: PR00685; TIFACTOR1B.  
 DR PROSITE: PS00782; TFIIB; 2.  
 KW Transcription regulation; Repeat; Zinc-finger.  
 FT ZN.FING  
 FT SEQUENCE 300 AA; 34097 MW; DE9758F398BC855F CRC64;

Query Match 31.4%; Score 38; DB 1; Length 300;  
 Best Local Similarity 44.4%; Pred. No. 14;  
 Matches 8; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 12 KXEEAVRLXXXXKNG 29  
 DB 127 KXVEEARLYREAVRNG 144

RESULT 9  
 NDPL\_MESAU STANDARD; PRT; 355 AA.  
 AC 060430;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE NEUROGENIC DIFFERENTIATION FACTOR 1 (BETA-CELL E-BOX TRANS-ACTIVATOR  
 DE 2) (BETA2).  
 GN NEUROD1 OR NEUROD.  
 OS Mesocricetus auratus (Golden hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Mesocricetus.  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-95293222; PubMed-7774807;  
 RA Naya F.J., Stellrecht C.M.M., Tsai M.-J.;  
 RT "Tissue-specific regulation of the insulin gene by a novel basic  
 RT helix-loop-helix transcription factor.";  
 RL Genes Dev. 9:1009-1019(1995).  
 CC -1- FUNCTION: ACTS AS A DIFFERENTIATION FACTOR DURING NEUROGENESIS.  
 CC TRANSCRIPTIONAL ACTIVATOR. BINDS TO THE INSULIN GENE E-BOX.  
 CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN. HETERODIMER WITH E47.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
 CC -1- TISSUE SPECIFICITY: MOST ABUNDANT IN PANCREATIC ALPHA- AND BETA-  
 CC CELLS, LESS IN BRAIN AND INTESTINE.

-1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 TRANSCRIPTION FACTORS. "ATONAL" SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: U24679; AA06518.1; ALT\_INIT.  
 DR HSSP: P10085; LMDY.  
 DR INTERPRO: IPR001092; -  
 DR INTERPRO: IPR003015; -  
 DR PFAM: PF00010; HLH; 1.  
 DR PROSITE: PS00038; HELIX\_LOOP\_HELIX; 1.  
 KW DNA-binding; Nuclear protein; Transcription regulation; Activator;  
 KW Neurogenesis; Developmental protein; Differentiation.  
 FT DOMAIN 58 77  
 FT DOMAIN 86 92  
 FT DNA\_BIND 101 112  
 FT DOMAIN 113 133  
 FT DOMAIN 67 75  
 FT DOMAIN 86 89  
 FT SEQUENCE 355 AA; 39763 MW; F4344D9D360226B2 CRC64;

Query Match 31.4%; Score 38; DB 1; Length 355;  
 Best Local Similarity 42.1%; Pred. No. 16;  
 Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 12 KXEEAVRLXXXXKNG 30  
 DB 39 KEDELEPMNAEDSLRNG 57

RESULT 10  
 NDPL\_MOUSE STANDARD; PRT; 357 AA.  
 AC 060867; Q60897;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE NEUROGENIC DIFFERENTIATION FACTOR 1.  
 GN NEUROD1 OR NEUROD.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX STRAIN-MFL AND 129/SV;  
 RX MEDLINE-95273957; PubMed-7754368;  
 RA Lee J.E., Hollenberg S.M., Snider L., Turner D.L., Lipnick N.,  
 RA Weintraub H.;  
 RT "Conversion of Xenopus ectoderm into neurons by Neurod, a basic  
 RT helix-loop-helix protein.";  
 RL Science 268:836-844(1995).  
 CC -1- FUNCTION: ACTS AS A DIFFERENTIATION FACTOR DURING NEUROGENESIS.  
 CC TRANSCRIPTIONAL ACTIVATOR. BINDS TO THE INSULIN GENE E-BOX.  
 CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN. HETERODIMER WITH E47.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN DIFFERENTIATING NEURONS OF  
 CC BOTH THE CENTRAL AND PERIPHERAL NERVOUS SYSTEMS.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING EMBRYONIC DEVELOPMENT.  
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS. "ATONAL" SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

DR EMBL: U28068; AAC52203.1; -  
 DR EMBL: U28888; AAC52204.1; -  
 DR HSSP: P10085; 1MDY.  
 DR MGD: MGI:1339708; NEUROD1.  
 DR INTERPRO: IPR001092; -  
 DR INTERPRO: IPR003015; -  
 DR PFAM: PF00010; HLH; 1.  
 DR PROSITE: PS00038; HELIX LOOP-HELIX; 1.  
 KW DNA-binding; Nuclear protein; transcription regulation; Activator;  
 KM Neurogenesis; Developmental protein; differentiation.  
 FT DOMAIN 58 77 GLU-RICH (ACIDIC).  
 FT DOMAIN 87 93 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DNA\_BIND 102 113 BASIC DOMAIN.  
 FT DOMAIN 114 154 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
 FT DOMAIN 58 64 POLY-GLU.  
 FT DOMAIN 67 77 POLY-GLU.  
 FT DOMAIN 87 90 POLY-LYS.  
 SO SEQUENCE 357 AA; 39998 MW; B6626E1315E31027 CRC64;

Query Match 31.4%; Score 38; DB 1; Length 357;  
 Best Local Similarity 42.1%; Pred. No. 16;  
 Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 12 KOXEEAVRLXXXXLKNGG 30  
 | : | | | : | : | | |  
 Db 39 KEDELEAMNNEEDSLRNGG 57

RESULT 11  
 ID NDF1\_RAT STANDARD; PRT; 357 AA.  
 AC 064289;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE NEUROGENIC DIFFERENTIATION FACTOR 1 (BASIC HELIX-LOOP-HELIX FACTOR 1) (BHF-1).  
 OS NEUROD1 OR NEUROD.  
 GN Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-CEREBELLUM;  
 RX MEDLINE-96220182; PubMed-8660336;  
 RA Kawakami H., Maruyama H., Yasunami M., Ohkubo H., Hara H., Salda T., Nakamishi S., Nakamura S.,  
 RT "Cloning and expression of a rat brain basic helix-loop-helix factor."  
 RT Blochem. Biophys. Res. Commun. 221:199-204(1996).  
 RL [2]  
 RP SEQUENCE OF 88-200 FROM N.A.  
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE-RETINA;  
 RA Ahmad I., Achary H.R.;  
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: ACTS AS A DIFFERENTIATION FACTOR DURING NEUROGENESIS.  
 CC TRANSCRIPTIONAL ACTIVATOR. BINDS TO THE INSULIN GENE E-BOX.  
 CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER BHLH PROTEIN. HETERODIMER WITH E47.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF TRANSCRIPTION FACTORS. "ATONAL" SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

DR EMBL: D82075; BAA11536.1; -  
 DR EMBL: D82074; BAA11535.1; -  
 DR EMBL: U80603; AAB38744.1; -  
 DR HSSP: P10085; 1MDY.  
 DR INTERPRO: IPR001092; -  
 DR INTERPRO: IPR003015; -  
 DR PFAM: PF00010; HLH; 1.  
 DR PROSITE: PS00038; HELIX-LOOP-HELIX; 1.  
 KW DNA-binding; Nuclear protein; transcription regulation; Activator;  
 KM Neurogenesis; Developmental protein; differentiation.  
 FT DOMAIN 58 77 GLU-RICH (ACIDIC).  
 FT DOMAIN 87 93 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DNA\_BIND 102 113 BASIC DOMAIN.  
 FT DOMAIN 114 154 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
 FT DOMAIN 67 76 POLY-GLU.  
 FT DOMAIN 87 90 POLY-LYS.  
 SO SEQUENCE 357 AA; 40000 MW; F773637E64D3E99E CRC64;

Query Match 31.4%; Score 38; DB 1; Length 357;  
 Best Local Similarity 42.1%; Pred. No. 16;  
 Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 12 KOXEEAVRLXXXXLKNGG 30  
 | : | | | : | : | | |  
 Db 39 KEDELEAMNNEEDSLRNGG 57

RESULT 12  
 ID NDF1\_ASEFM2 STANDARD; PRT; 419 AA.  
 AC P26813;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE DNA LIGASE (EC 6.5.1.1) (POLYDEOXYRIBONUCLEOTIDE SYNTHASE [ATP]).  
 OS African swine fever virus (isolate Malawi 11/20/1) (ASFV).  
 OC Viruses; dsDNA viruses, no RNA stage;  
 OC African swine fever-like viruses.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-92310959; PubMed-1614852;  
 RA Hammond J.M., Kerr S.M., Smith G.L., Dixon L.K.;  
 RT "An African swine fever virus gene with homology to DNA ligases."  
 RL Nucleic Acids Res. 20:2667-2671(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-94292916; PubMed-8021596;  
 RA Dixon L.K., Twigg S.R.F., Baylis S.A., Vydelingum S., Bristol C., Hammond J.M., Smith G.L.;  
 RT "Nucleotide sequence of a 55 kbp region from the right end of the genome of a pathogenic African swine fever virus isolate (Malawi 11/20/1)."  
 RT J. Gen. Virol. 75:1655-1684(1994).  
 RL J. Gen. Virol. 75:1655-1684(1994).  
 CC -1- FUNCTION: THIS PROTEIN SEALS DURING DNA REPLICATION, DNA RECOMBINATION AND DNA REPAIR NICKS IN DOUBLE-STRANDED DNA.  
 CC IT IS NOT ESSENTIAL FOR VIRAL REPLICATION AND RECOMBINATION.  
 CC -1- CATALYTIC ACTIVITY: ATP + (DEOXYRIBONUCLEOTIDE) (N) + (DEOXYRIBONUCLEOTIDE) (N+M) -> AMP + PYROPHOSPHATE + (DEOXYRIBONUCLEOTIDE) (N+M).  
 CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT DNA LIGASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X65192; CAA6310.1; -  
 DR EMBL: X71982; CAA50805.1; -  
 DR PIR: S23618; S23018.  
 DR INTERPRO: IPR000977; -  
 DR PFAM: PF01068; DNA\_Ligase; 1.  
 DR PROSITE: PS00697; DNA\_LIGASE\_A1; 1.  
 DR PROSITE: PS00333; DNA\_LIGASE\_A2; 1.  
 DR PROSITE: PS00160; DNA\_LIGASE\_A3; 1.  
 DR DNA repair; DNA replication; DNA recombination; Cell division; Ligase;  
 KW ATP-binding.  
 FT BINDING 151 151 AMP (BY SIMILARITY).  
 SO SEQUENCE 419 AA; 48041 MW; DA781C64CA1B10F0 CRC64;

Query Match 31.4%; Score 38; DB 1; Length 419;  
 Best Local Similarity 42.1%; Pred. No. 19;  
 Matches 8; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 11 SKOXEEAVRLXXXXKNG 29  
 ID 271 SKNDEALRLKTOFKEG 289

RESULT 13  
 L01A\_LYCEN STANDARD; PRT; 589 AA.  
 AC 004973;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE 2-ISOPROPYLMALATE SYNTHASE A (EC 4.1.3.12) (ALPHA-ISOPROPYLMALATE  
 SYNTHASE A) (ALPHA-IPM SYNTHASE A).  
 GN IPMSA.  
 OS Lycopersicon pennellii (Tomato).  
 CC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 CC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;  
 CC Solanales; Solanaceae; Solanum.  
 CC [1]  
 RN SEQUENCE FROM N.A.  
 RC TISSUE-LEAF;  
 RA Wei T., Maite D., Steffens J.C.;  
 RT "Cloning of two L. pennellii 2-isopropylmalate synthase cDNA and  
 RT their functional expression in yeast."  
 RL Submitted (MAY-1997) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: CATALYZES CONDENSATION OF ACETYL-COA AND 2-  
 CC OXOISOBALATE TO FORM 2-ISOPROPYLMALATE SYNTHASE.  
 CC -1- CATALYTIC ACTIVITY: 3-CARBOXY-3-HYDROXY-4-METHYLENTRANOATE + COA -  
 CC ACETYL-COA + 3-METHYL-2-OXOBUTANOATE + H(2)O.  
 CC -1- PATHWAY: FIRST STEP IN LEOICINE BIOSYNTHESIS.  
 CC -1- SIMILARITY: BELONGS TO THE ALPHA-IPM SYNTHETASE / HOMOCITRATE  
 CC SYNTHASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: AF004165; AAB61598.1; -  
 DR HSSP: P27320; IDOY.  
 DR INTERPRO: IPR000891; -  
 DR INTERPRO: IPR002034; -  
 DR PFAM: PF00682; HMG-Like; 1.  
 DR PROSITE: PS00815; AIPM\_HOMOCIT\_SYNTH\_1; 1.  
 DR PROSITE: PS00816; AIPM\_HOMOCIT\_SYNTH\_2; FALSE\_NEG.  
 KW Leucine biosynthesis; Lyase.  
 SO SEQUENCE 589 AA; 64360 MW; 150E48900188BDDF CRC64;

Query Match 31.4%; Score 38; DB 1; Length 589;  
 Best Local Similarity 36.0%; Pred. No. 27;

Matches 9; Conservative 5; Mismatches 11; Indels 0; Gaps 0;  
 OY 11 SKOXEEAVRLXXXXKNGXSSGA 35  
 ID 40 SKIDPKYVRIFDTLNDGQSPGA 64

RESULT 14  
 AS2\_YEAST STANDARD; PRT; 430 AA.  
 AC P39945;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE AS22 PROTEIN.  
 GN AS22 OR YER101C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 CC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;  
 CC Saccharomycetaceae; Saccharomyces.  
 CC [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-S288C / AB972;  
 RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,  
 RA Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,  
 RA Chung E., Duncan M., Guzman E., Hartzell G., Huntcke-Smith S.,  
 RA Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,  
 RA Moseedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,  
 RA Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,  
 RA Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.,  
 RL Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.  
 CC -1- SIMILARITY: HIGH, TO YEAST AST1; ALSO SIMILAR TO YMR152W.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: U18839; AAB64656.1; -  
 DR SGD: S0000903; AST2.  
 DR SGD: S0000903; AST2.  
 SO SEQUENCE 430 AA; 48370 MW; FBEC2F2CDEB94F84 CRC64;

Query Match 30.6%; Score 37; DB 1; Length 430;  
 Best Local Similarity 30.0%; Pred. No. 30;  
 Matches 9; Conservative 5; Mismatches 16; Indels 0; Gaps 0;

OY 4 GTXXXXSKOXEEAVRLXXXXKNGXSS 33  
 ID 182 GTALNLAQLEKNDLNTESNVLLNGRGS 211

RESULT 15  
 YKTL\_CAEEL STANDARD; PRT; 488 AA.  
 ID YKTL\_CAEEL  
 AC P34312;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-JUN-1994 (Rel. 29, Last annotation update)  
 DE HYPOTHETICAL 54.7 KDA PROTEIN C07A9.1 IN CHROMOSOME III.  
 GN C07A9.1.  
 OS Caenorhabditis elegans.  
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 CC Rhabditidae; Peloderinae; Caenorhabditis.  
 CC [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE-94150718; PubMed-7906398;  
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,

RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,  
 RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,  
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,  
 RA Sims M., Smaildon N., Smith A., Smith M., Sonhammer E., Staden R.,  
 RA Sulston J., Thelery-Mieg J., Thomas K., Vaudin M., Vaughan K.,  
 RA Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,  
 RA Wohldman P.,  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 elegans.";  
 RL Nature 368:32-38(1994).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z29094; CAA82340.1; -.  
 DR PIR; S40706; S40706.  
 DR HSSP; P23807; 11XX.  
 DR WORMPEP; C07A9.1; CE00502.  
 DR INTERPRO; IPR001304; -.  
 DR PFAM; PF00059; lectin\_c; 1.  
 DR PROSITE; PS50041; C\_TYPE\_LLECTIN\_2; UNKNOWN\_1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 488 AA; 54717 MW; 9022691E47078814 CRC64;

Query Match 30.6%; Score 37; DB 1; Length 488;  
 Best Local Similarity 30.0%; Pred. No. 34;  
 Matches 9; Conservative 3; Mismatches 18; Indels 0; Gaps 0;

QY 5 TXXXXXKQXEEAVRLXXXXLKNGXSSG 34  
 | : | | : | | | : |  
 DB 445 TEMSRSRKEKETEDSINVKSLEKGGTARG 474

Search completed: February 13, 2001, 16:09:51  
 Job time: 187 sec



```

OC Alpharpesvirinae; Simplexvirus.
OX NCBI_TaxID=10310;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HG52;
RX MEDLINE=87111457; PubMed=3027242;
RA McGeoch D.J., Moss H.W., McNab D., Frame M.C.;
RT "DNA sequence and genetic content of the HindIII 1 region in the short
RT unique component of the herpes simplex virus type 2 genome:
RT identification of the gene encoding glycoprotein G, and evolutionary
RT comparisons."
RN J. Gen. Virol. 68:19-38(1987).
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN-HG52;
RX MEDLINE=90278430; PubMed=2161906;
RA Everett R., Fenwick M.;
RT "Comparative DNA sequence analysis of the host shutoff genes of
RT different strains of herpes simplex virus: type 2 strain HG52 encodes
RT a truncated UL41 product."
RN J. Gen. Virol. 71:1387-1390(1990).
RL [3]
RP SEQUENCE FROM N.A.
RC STRAIN-HG52;
RX MEDLINE=92113549; PubMed=1662697;
RA McGeoch D.J., Cunningham C., McIntyre G., Dolan A.;
RT "Comparative sequence analysis of the long repeat regions and
RT adjoining parts of the long unique regions in the genomes of herpes
RT simplex viruses types 1 and 2."
RN J. Gen. Virol. 72:3057-3075(1991).
RL [4]
RP SEQUENCE FROM N.A.
RC STRAIN-HG52;
RX MEDLINE=92356101; PubMed=1332965;
RA Barnett B.C., Dolan A., Telford E.A.R., Davison A.J., McGeoch D.J.;
RT "A novel herpes simplex virus gene (UL49A) encodes a putative membrane
RT protein with counterparts in other herpesviruses."
RN J. Gen. Virol. 73:2167-2171(1992).
RL [5]
RP SEQUENCE FROM N.A.
RC STRAIN-HG52;
RX Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RA EMBL: 286099; CAB06753.1; -.
DR INTERPRO: IPR000501; -.
DR PIRAM: PR01366; PRTP; 1.
SQ SEQUENCE 785 AA; 85240 MW; 246988E41997DF62 CRC64;

```

Query Match 36.4%; Score 44; DB 12; Length 785;  
 Best Local Similarity 41.7%; Pred. No. 14;  
 Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

```

OY 12 KQXEEAVRLXXXXLKNKGXSGA 35
Db 422 EOCDEALRVRLRLGAGATGA 445

```

```

RESULT 3
O9Y717 PRELIMINARY; PRT; 141 AA.
AC O9Y717;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DE 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE CONSERVED HYPOTHETICAL PROTEIN (FRAGMENT).
GN SPAC8E11.11.
OS Schizosaccharomyces pombe (Fission Yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
OC Schizosaccharomycetaceae; Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-972H-;
RA McLean J., Harris D.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-972H-;
RX Wood V., Barrell B.G., Rajandream M.A.;
RA Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL021817; CAB40198.1; -.
DR INTERPRO: IPR002110; -.
FT NON_TER 141
SQ SEQUENCE 141 AA; 15804 MW; BF8BFD97A361EC4F CRC64;

```

Query Match 33.9%; Score 41; DB 3; Length 141;  
 Best Local Similarity 34.8%; Pred. No. 7.2;  
 Matches 8; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

```

OY 12 KQXEEAVRLXXXXLKNKGXSG 34
Db 64 KEEVOAIEVTKILSGGWNG 86

```

```

RESULT 4
O9RRJ0 PRELIMINARY; PRT; 157 AA.
AC O9RRJ0;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE HYPOTHETICAL 17.0 KDA PROTEIN.
GN DR2500.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RL;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodsan R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J., Lam P., McDonald L., Uetebach T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome Sequence of the Radioresistant Bacterium Deinococcus
RT radiodurans RI."
RL Science 286:1571-1577(1999).
DR EMBL: AE002079; AAF12045.1; -.
DR TIGR: DR2500; -.
KW Hypothetical protein.
SQ SEQUENCE 157 AA; 17027 MW; B76BD89F60A5B5D CRC64;

```

Query Match 33.9%; Score 41; DB 2; Length 157;  
 Best Local Similarity 42.1%; Pred. No. 8.1;  
 Matches 8; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

```

OY 16 EEAVALXXXXLKNKGXSG 34
Db 74 DDAVQVFRALKNAGLDSG 92

```

```

RESULT 5
O42143 PRELIMINARY; PRT; 266 AA.
AC O42143;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE PROGLUCAGON I.
OS Xenopus laevis (African clawed frog).

```



OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
OC Xenopodidae; Xenopus.  
OX NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=97368292; PubMed=9223287;  
RA Irwin D.M., Satkunarajah M., Wen Y., Brubaker P.L., Pederson R.A.,  
RA Wheeler M.B.;  
RT "The Xenopus produgacgon gene encodes novel GLP-1-like peptides with  
RT insulinotropic properties."  
RL Proc. Natl. Acad. Sci. U.S.A. 94:7915-7920(1997).  
DR EMBL; AF04432; AAB5660.1; -.  
DR HSSP; P01274; IGCN.  
DR INTERPRO; IPR000532; -.  
DR PFAM; PF00123; hormone2; 5.  
DR PRINTS; PR00275; GLUCAGON; 5.  
DR PROSITE; PS00280; GLUCAGON; 5.  
DR PRODOM; PD002392; -. 1.  
DR PRODOM; PD002700; -. 3.  
SQ SEQUENCE 266 AA; 30951 MW; 544F7BCC20AF872C CRC64;

Query Match 33.9%; Score 41; DB 13; Length 266;  
Best Local Similarity 34.5%; Pred. No. 15;  
Matches 10; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

OY 4 GTXXXXXKXEEAVRLXXXXLKNGXS 32  
DB 100 GTTSDVTQQLDEKAKFIDWLINGSPS 128

RESULT 6  
ID 092527 PRELIMINARY; PRT; 306 AA.  
AC 092527;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
DE COAT PROTEIN (CAPSID PROTEIN).  
OS Carnation latent virus (CLV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.  
OX NCBI\_TaxID=12164;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Meenan B.M.;  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RX SEQUENCE FROM N.A.  
RX MEDLINE=91324119; PubMed=1713905;  
RA Meenan B.M., Mills P.R.;  
RT "Nucleotide sequence of the 3'-terminal region of carnation latent  
RT virus."  
RL Intervirology 32:262-267(1991).  
DR EMBL; AJ010697; CA09306.1; -.  
DR INTERPRO; IPR000052; -.  
DR PFAM; PF00286; virus\_P-coat; 1.  
DR PRINTS; PR00232; POTXCARCOAT.  
DR PROSITE; PS00418; POTEX\_CARLAVIRUS\_COAT; 1.  
DR PRODOM; PD000603; -. 1.  
KW Coat protein.  
SQ SEQUENCE 306 AA; 33890 MW; 4456EBB53E174298 CRC64;

Query Match 33.9%; Score 41; DB 12; Length 306;  
Best Local Similarity 43.5%; Pred. No. 17;  
Matches 10; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

OY 12 KXEEAVRLXXXXLKNGXS 34  
DB 62 KLFNFNSONTAGELKNGFESG 84

RESULT 7  
ID 09UYT6 PRELIMINARY; PRT; 402 AA.  
AC 09UYT6;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
DE MOLYBDENUM COPACFOR BIOSYNTHESIS PROTEIN (MOEA-1).  
GN PA31436.  
OS Pyrococcus abyssi.  
OC Archaea; Euryarchaeota; Thermococcales; Pyrococcus.  
OX NCBI\_TaxID=29292;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ORSAY;  
RA Hellig R.;  
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome  
RT structure and evolution."  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ248287; CAB50326.1; -.  
DR INTERPRO; IPR001453; -.  
DR INTERPRO; IPR002106; -.  
DR PFAM; PF00994; MOCF\_biosynth; 1.  
DR PROSITE; PS00339; AA\_tRNA\_LIGASE\_II\_2; UNKNOWN\_1.  
SQ SEQUENCE 402 AA; 43327 MW; 44545EDA70F6A78E CRC64;

Query Match 33.9%; Score 41; DB 1; Length 402;  
Best Local Similarity 39.1%; Pred. No. 23;  
Matches 9; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

OY 12 KXEEAVRLXXXXLKNGXS 34  
DB 237 KELIEGVAVADIVYISGASGS 259

RESULT 8  
ID 075464 PRELIMINARY; PRT; 589 AA.  
AC 075464;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
DE NUCLEAR MATRIX PROTEIN NRP/B.  
GN NRPB.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98234394; PubMed=9566959;  
RA Kim T.-A., Lim J., Oca S., Raja S., Rogers R., Rivnay B., Avraham H.,  
RA Avraham S.;  
RT "NRP/B, a novel nuclear matrix protein, associates with p110(RB) and  
RT is involved in neuronal differentiation."  
RL J. Cell Biol. 141:553-566(1998).  
DR EMBL; AF059611; AAC26109.1; -.  
DR INTERPRO; IPR000169; -.  
DR INTERPRO; IPR000210; -.  
DR INTERPRO; IPR001798; -.  
DR PFAM; PF00651; BTB; 1.  
DR PFAM; PF01344; Kelch; 6.  
DR PROSITE; PS00639; THIOL\_PROTEASE\_HTS; UNKNOWN\_1.  
KW Matrix protein.  
SQ SEQUENCE 589 AA; 66129 MW; DB003A1DFA65BAA0 CRC64;

Query Match 33.9%; Score 41; DB 4; Length 589;  
Best Local Similarity 45.0%; Pred. No. 36;  
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

OY 11 SKXEEAVRLXXXXLKNG 30

```

Db      262 SKEIVEAIRCKLKILQNDG 281

RESULT 9
Q9PBG9 PRELIMINARY; PRT; 589 AA.
AC Q9PBG9
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
DE ECTODERM-NEURAL CORTEX-1 PROTEIN.
EN ENC-1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
ON NCBI_TaxID=9606;
RX MEDLINE=98350113; PubMed=9683534;
RA Hernandez M.-C., Andres-Barguin P.J., Holt I., Israel M.A.;
RT "Cloning of human ENC-1 and evaluation of its expression and
RT regulation in nervous system tumors.";
RL EMBL: AF005381; AAC64498.1; -.
DR INTERPRO: IPR000169; -.
DR INTERPRO: IPR000210; -.
DR INTERPRO: IPR001798; -.
DR PFAM: PF00651; BTB; 1.
DR PRAM: PF01344; Kelch; 6.
DR PROSITE: PS00639; THIOLEPROTEASE_HIS; UNKNOWN_1.
SQ SEQUENCE 589 AA; 66113 MW; C0002116A6EALAB CRC64;

```

```

Query Match 33.9%; Score 41; DB 4; Length 589;
Best local Similarity 45.0%; Pred. No. 36;
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;
QY 11 SKOXEEFVRLXXXXKNGG 30
DB 262 SKEIVEAIRCKLKILQNDG 281

```

```

RESULT 10
ID Q35709 PRELIMINARY; PRT; 589 AA.
AC Q35709
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE ECTODERM-NEURAL CORTEX-1 PROTEIN (ENC-1).
EN ENC1 OR ENC-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
ON NCBI_TaxID=10090;
RX MEDLINE=97252647; PubMed=9096139;
RA Hernandez M.-C., Andres-Barguin P.J., Martinez S., Bulfone A.,
RA Rubenstein J.L.R., Israel M.A.;
RT "ENC-1: a novel mammalian kelch-related gene specifically expressed in
RT the nervous system encodes an actin-binding protein.";
RL J. Neurosci. 17:3038-3051(1997).
DE FUNCTION: ACTIN-BINDING PROTEIN INVOLVED IN THE REGULATION OF
DE NEURAL PROGENITOR FORMATION AND IN DIFFERENTIATION OF NEURAL CREST
DE CELLS.
CC -1- TISSUE SPECIFICITY: PRIMARILY EXPRESSED IN THE NERVOUS SYSTEM.
CC -1- DEVELOPMENTAL STAGE: EXPRESSION IS HIGHLY DYNAMIC BUT MOSTLY
CC RESTRICTED TO THE NS. OUTSIDE THE NS, EXPRESSION IS DETECTED IN
CC THE ROSTRAL-MOST SOMITOMERE OF THE PRESMOTIC MESODERM, AT THE
CC TIMES CORRESPONDING TO THE EPITHELIALIZATION THAT PRECEDES SOMITE

```

```

CC FORMATION. FIRST DETECTED IN THE BRAIN AND SPINAL CHORD OF 12 PC
CC EMBRYOS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. INTERACTS WITH THE ACTIN
CC CYTOSKELETON.
CC -1- SIMILARITY: BELONGS TO THE KELCH/MLP FAMILY.
CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
DR EMBL: U65079; AAB64206.1; -.
DR MGD: MG1:109610; Enc1.
DR INTERPRO: IPR000210; -.
DR INTERPRO: IPR001798; -.
DR PFAM: PF00651; BTB; 1.
DR PFAM: PF01344; Kelch; 6.
KW Phosphorylation; Actin-binding; Developmental protein; Cytoskeleton.
FT DOMAIN 28; 144
FT REPEAT 296; 340
FT REPEAT 341; 388
FT REPEAT 389; 444
FT REPEAT 445; 492
FT REPEAT 493; 538
FT REPEAT 539; 585
SQ SEQUENCE 589 AA; 66085 MW; 12E62354D508B6A2 CRC64;

```

```

Query Match 33.9%; Score 41; DB 11; Length 589;
Best local Similarity 45.0%; Pred. No. 36;
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;
QY 11 SKOXEEFVRLXXXXKNGG 30
DB 262 SKEIVEAIRCKLKILQNDG 281

```

```

RESULT 11
ID Q9VRN7 PRELIMINARY; PRT; 2630 AA.
AC Q9VRN7
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE CG5256 PROTEIN.
EN CG5256.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
ON NCBI_TaxID=7227;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceolinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amandites P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertan B.P., Bhandari D., Bolintsov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegam C.,
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lai Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

```

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman K.A., Pan S., Pollard J., Puti V., Reese M.G.,  
 RA Relbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spler E., Spredling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhang G., Zhao Q., Zheng L.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003565; AAF50755.1; -.  
 DR HSSP: P08567; 1PLS.  
 DR FLYBASE: FBgn0035628; CG5256.  
 DR INTERPRO: IPR000219; -.  
 DR INTERPRO: IPR001331; -.  
 DR INTERPRO: IPR001478; -.  
 DR INTERPRO: IPR001849; -.  
 DR INTERPRO: IPR003116; -.  
 DR PFAM: PF00169; PH: 2.  
 DR PFAM: PF00621; RhogEF. 1.  
 DR PFAM: PF02196; RBD. 1.  
 DR PROSITE: PS00741; GDS\_CDC24; UNKNOWN\_1.  
 DR PROSITE: PS50003; PH\_DOMAIN. 1.  
 SQ SEQUENCE 2630 AA; 292511 MW; 56D1AD45971FE5B9 CRC64;

Query Match 33.9%; Score 41; DB 5; Length 2630;  
 Best Local Similarity 41.7%; Pred. No. 1.9e+02;  
 Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;  
 QY 12 KOXEEAVRLXXXXLKNGXSSGA 35  
 Db 1762 RQIRRESVRNMSIPMKNFSGSS 1785

RESULT 12  
 ID 09VRN8 PRELIMINARY; PRT; 2637 AA.  
 AC 09VRN8:  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE CG5406 PROTEIN.  
 GN CG5406.  
 OS *Drosophila melanogaster* (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.H.C., Blazer V.G., Champe M., Pfeiffer B.D.,  
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Mikos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glosker A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harrey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.U., Wei M.-H., Ibegwan C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman K.A., Pan S., Pollard J., Puti V., Reese M.G.,  
 RA Relbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spler E., Spredling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhang G., Zhao Q., Zheng L.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003565; AAF50755.1; -.  
 DR HSSP: P08567; 1PLS.  
 DR FLYBASE: FBgn0035629; CG5406.  
 DR INTERPRO: IPR000219; -.  
 DR INTERPRO: IPR001331; -.  
 DR INTERPRO: IPR001478; -.  
 DR INTERPRO: IPR001849; -.  
 DR INTERPRO: IPR003116; -.  
 DR PFAM: PF00169; PH: 2.  
 DR PFAM: PF00621; RhogEF. 1.  
 DR PFAM: PF02196; RBD. 1.  
 DR PROSITE: PS00741; GDS\_CDC24; UNKNOWN\_1.  
 DR PROSITE: PS50003; PH\_DOMAIN. 1.  
 SQ SEQUENCE 2637 AA; 293079 MW; 76F475BA3CBE961B CRC64;

Query Match 33.9%; Score 41; DB 5; Length 2637;  
 Best Local Similarity 41.7%; Pred. No. 1.9e+02;  
 Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;  
 QY 12 KOXEEAVRLXXXXLKNGXSSGA 35  
 Db 1769 RQIRRESVRNMSIPMKNFSGSS 1792

RESULT 13  
 ID P96631 PRELIMINARY; PRT; 127 AA.  
 AC P96631:  
 DT 01-MAY-1997 (TREMBlrel. 03, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)  
 DE PROBABLE REPRESSOR PROTEIN.  
 GN YDCN.  
 OS *Bacillus subtilis*.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=1423;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98000887; PubMed=9341680;  
 RA Beloin C., Ayora S., Exley R., Hirschbein L., Ogasawara N.,  
 RA Katsuhara Y., Alonso J.C., Le Hegarat F.;  
 RT "Characterization of an *lfp*-like (*lfpC*) gene from *Bacillus subtilis*.";  
 RL Mol. Gen. genet. 256:63-71(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;

RA MEDLINE-98044033; PubMed-9384377;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertero M.G., Bessler P., Bolotin A., Borchart S.,  
 RA Brouillet S., Brouillet L., Brans A., Brun M., Brynne S.C., Bron S.,  
 RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Carter N.M.,  
 RA Denton F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Frits C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N.,  
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 RA Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasanara Y., Kiberr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,  
 RA Pario V., Pohl T.M., Portetelle D., Potwilk S., Prescott A.M.,  
 RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,  
 RA Sato T., Scanlan E., Schleich S., Schoeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Serron P., Shin B.S., Soldo B.,  
 RA Sorokin A., Taccini E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandendol M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Welternegger T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,  
 RT "The complete genome sequence of the gram-positive bacterium *Bacillus*  
 RT subtilis."

RT Nature 390:249-256(1997).  
 RL Nature 390:249-256(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-168:  
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;  
 RL Submitted (Nov-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB001488; BAI19320.1; -  
 DR EMBL: 299106; CAB12289.1; -  
 DR INTERPRO: IPR001387; -  
 DR PIRAM: PF01381; HTH\_3; 1  
 SQ SEQUENCE 127 AA; 14649 MW; 3CC91D5B1D51628C CRC64;

Query Match 33.1%; Score 40; DB 2; Length 127;  
 Best Local Similarity 47.1%; Pred. No. 9.7;  
 Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 13 QXEEAVRLXXXXKNG 29  
 Db 100 EFDEETARLVKALKNG 116

RESULT 14  
 O90184 PRELIMINARY; PRT; 374 AA.  
 AC O90184;  
 DT 01-MAY-2000 (TREMREL. 13, Created)  
 DT 01-MAY-2000 (TREMREL. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMREL. 13, Last annotation update)  
 DE HYPOTHEICAL 41.7 KDA PROTEIN.  
 GN I3238.06.  
 OS Leishmania major.  
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBI\_TaxID=5664;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-FRIEDLIN;  
 RA Wedler H., Hilbert H., Dusterhoft A., Ivens A.C., Murphy L.,  
 RA Quail M., Rajandream M.A., Barrell B.G.;  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.

RC STRAIN-FRIEDLIN;  
 RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,  
 RA Smith D.F.;  
 RT "A physical map of the *Leishmania major* Friedlin genome.";  
 RL Genome Res. 8:135-145(1998).  
 DR EMBL: AL133466; CAB63133.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 374 AA; 41675 MW; AA38847DBE433937 CRC64;

Query Match 33.1%; Score 40; DB 5; Length 374;  
 Best Local Similarity 39.1%; Pred. No. 33;  
 Matches 9; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

OY 12 KQXEEAVRLXXXXKNGXSSG 34  
 Db 14 KQKKEAKKASQGPLRGAGSSG 36

RESULT 15  
 O9LHL3 PRELIMINARY; PRT; 455 AA.  
 AC O9LHL3;  
 DT 01-OCT-2000 (TREMREL. 15, Created)  
 DT 01-OCT-2000 (TREMREL. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMREL. 15, Last annotation update)  
 DE RNA-BINDING PROTEIN-LIKE.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-COLUMBIA;  
 RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;  
 RL Submitted (May-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-COLUMBIA;  
 RA Nakamura Y.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.  
 RT Sequence features of the regions of 4,251,695 bp covered by ninety pl,  
 RT TAC and BAC clones."  
 RL DNA Res. 7:217-221(2000).  
 DR EMBL: AP002041; BAB02607.1; -  
 SQ SEQUENCE 455 AA; 48958 MW; E88117B22C33BBA9 CRC64;

Query Match 33.1%; Score 40; DB 10; Length 455;  
 Best Local Similarity 41.7%; Pred. No. 41;  
 Matches 10; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

OY 12 KQXEEAVRLXXXXKNGXSSG 35  
 Db 5 QREVESVTEKRRLDGGSSGA 28

Search completed: February 13, 2001, 16:07:28  
 Job time: 83 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2001, 16:06:05 ; Search time 36.73 Seconds  
(without alignments)  
127.643 Million cell updates/sec

Title: US-08-908-867-38

Perfect score: 121  
Sequence: 1 XXXGTXXXXXXKQXEEAVRLXXXLKNCGXSSGAXXXXX 40

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPTRMBL\_15:\*

2: sp\_archaea:\*

3: sp\_bacteria:\*

4: sp\_fungi:\*

5: sp\_human:\*

6: sp\_invertebrate:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phase:\*

10: sp\_plant:\*

11: sp\_proteint:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	36.4	234	5	Q9NM02
2	44	36.4	785	12	P89451
3	41	33.9	141	3	Q9Y717
4	41	33.9	157	2	Q9RR10
5	41	33.9	266	13	Q42143
6	41	33.9	306	12	O92527
7	41	33.9	402	1	O9UYT6
8	41	33.9	589	4	O75464
9	41	33.9	589	4	O9UPG9
10	41	33.9	589	11	O35709
11	41	33.9	2630	5	O9VRN7
12	41	33.9	2637	5	O9VRN8
13	40	33.1	127	2	P96631
14	40	33.1	374	5	O9U184
15	40	33.1	455	10	O9LHJ3
16	40	33.1	609	10	O9SD72
17	40	33.1	731	5	O9VZK7
18	40	33.1	766	5	O9N9B6
19	40	33.1	772	10	O9SN69

20	40	33.1	1296	2	O9KX3	O9KX3 mycoplasma
21	40	33.1	2382	5	O9NKP4	O9NKP4 leishmania
22	39	32.2	145	2	P70746	P70746 aeromonas h
23	39	32.2	208	1	O58594	O58594 pyrococcus
24	39	32.2	342	2	O92302	O92302 pseudomonas
25	39	32.2	342	2	O9R733	O9R733 pseudomonas
26	39	32.2	342	2	O9R277	O9R277 pseudomonas
27	39	32.2	343	2	O31180	O31180 pseudomonas
28	39	32.2	688	2	O25812	O25812 helicobacte
29	39	32.2	688	2	O92K11	O92K11 helicobacte
30	38.5	31.8	472	2	O9KX2	O9KX2 streptomyce
31	38.5	31.8	653	10	O41729	O41729 zea mays (m
32	38.5	31.8	1702	11	O54875	O54875 ratius norv
33	38	31.4	214	12	O9P206	O9P206 hepatitis d
34	38	31.4	239	10	O9LTV4	O9LTV4 arbidopsi
35	38	31.4	241	5	O04317	O04317 scapromyza
36	38	31.4	241	5	O99183	O99183 scapromyza
37	38	31.4	300	1	O9V0V5	O9V0V5 pyrococcus
38	38	31.4	421	2	O9KX47	O9KX47 bacterioph
39	38	31.4	421	9	O9XJW0	O9XJW0 bacterioph
40	38	31.4	540	10	O9L737	O9L737 arbidopsi
41	38	31.4	622	10	O9L1B8	O9L1B8 zea mays (m
42	38	31.4	636	10	O9SD69	O9SD69 arbidopsi
43	38	31.4	726	5	O9X1T8	O9X1T8 caenorhadi
44	38	31.4	1464	5	O96782	O96782 drosophila
45	38	31.4	1464	5	O9VX75	O9VX75 drosophila

#### ALIGNMENTS

RESULT	1	ALIGNMENTS
Q9NM02	PRELIMINARY;	PRT; 234 AA.
AC Q9NM02:		
DT 01-OCT-2000 (TRENBLREL.15, Created)		
DT 01-OCT-2000 (TRENBLREL.15, Last sequence update)		
DT 01-OCT-2000 (TRENBLREL.15, Last annotation update)		
DE POSSIBLE HYPOTHETICAL 45.5 KDA PROTEIN (FRAGMENT).		
GN LM26.290		
OS Leishmania major.		
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.		
OX NCBI_TaxID=5664;		
RN [1]		
RP SEQUENCE FROM N.A.		
RC STRAIN=FREDLIN;		
RA Murphy L., Quail M., Harris D., Rajandream M., Trems A., Barrell B.;		
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.		
DR EMBL; AL160493; CAB97908.1; -		
FT NON_TER	234	
FT SEQUENCE	234 AA; 24954 MW; 0F013FAB8A1196FA CRC64;	
Query Match	36.4%; Score 44; DB 5; Length 234;	
Best Local Similarity	44.4%; Pred. No. 3.6;	
Matches	12; Conservative 3; Mismatches 10; Indels 2; Gaps 1;	
QY	11 SKQXEEAV-RLXXXXLKNCGXSSGA 35	
DB	148 SKQVREKALAAWLSDALVNGAPSGA 174	
RESULT 2		
P89451	PRELIMINARY;	PRT; 785 AA.
AC P89451:		
DT 01-MAY-1997 (TRENBLREL.03, Created)		
DT 01-MAY-1997 (TRENBLREL.03, Last sequence update)		
DT 01-MAY-2000 (TRENBLREL.13, Last annotation update)		
DE HERPES SIMPLEX VIRUS TYPE 2 (STRAIN HG52), COMPLETE GENOME.		
GN UL28.		
OS Herpes simplex virus (type 2).		
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;		

```

OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10310;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HG52;
RX MEDLINE=87111457; PubMed=3027242;
RA McGeoch D.J., Moss H.W., McInyre G., Frame M.C.;
RT "DNA sequence and genetic content of the HindIII 1 region in the short
RT unique component of the herpes simplex virus type 2 genome:
RT identification of the gene encoding glycoprotein G, and evolutionary
RT comparisons."
RL J. Gen. Virol. 68:19-38(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-HG52;
RX MEDLINE=90278430; PubMed=2161906;
RA Everett R., Fenwick M.;
RT "Comparative DNA sequence analysis of the host shutoff genes of
RT different strains of herpes simplex virus: type 2 strain HG52 encodes
RT a truncated UL41 product."
RL J. Gen. Virol. 71:1387-1390(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-HG52;
RX MEDLINE=92113549; PubMed=1662697;
RA McGeoch D.J., Cunningham C., McIntyre G., Dolan A.;
RT "Comparative sequence analysis of the long repeat regions and
RT adjoining parts of the long unique regions in the genomes of herpes
RT simplex viruses types 1 and 2."
RL J. Gen. Virol. 72:3057-3075(1991).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-HG52;
RX MEDLINE=92356101; PubMed=1322965;
RA Barrett B.C., Dolan A., Telford E.A.R., Davison A.J., McGeoch D.J.;
RT "A novel herpes simplex virus gene (UL49a) encodes a putative membrane
RT protein with counterparts in other herpesviruses."
RL J. Gen. Virol. 73:2167-2171(1992).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN-HG52;
RA Dolan A.;
RT Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: 266059; CAB06753.1; -.
DR INTERPRO: IPR000501; -.
DR PFM: PF01366; PRP; 1.
SQ SEQUENCE 785 AA; 85240 MW; 246988E4197DF62 CRC64;

Query Match 36.4%; Score 44; DB 12; Length 785;
Best Local Similarity 41.7%; Pred. No. 14;
Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

OY 12 KOXEEAVRLXXXLKNKGXSSG 35
DB 422 EOCDEALRLVRLARLGAGATGCA 445

```

```

RC STRAIN-972H-;
RA McLean J., Harris D.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-972H-;
RA Wood V., Barrell B.G., Rajandream M.A.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL021817; CAB40198.1; -.
DR INTERPRO: IPR002110; -.
FT NON TER 141.
SQ SEQUENCE 141 AA; 15804 MW; BF8BF97A361EC4F CRC64;

Query Match 33.9%; Score 41; DB 3; Length 141;
Best Local Similarity 34.8%; Pred. No. 7.2;
Matches 8; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

OY 12 KOXEEAVRLXXXLKNKGXSSG 34
DB 64 KETEOVALEVTWILSLNGVWNG 86

RESULT 4
OQRRJ0 PRELIMINARY; PRT; 157 AA.
AC OQRRJ0;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE HYPOTHETICAL 17.0 KDA PROTEIN.
GN DR2500.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RI;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Shen M.,
RA Vanathavan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
RA Makarova K.S., Atavind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome Sequence of the Radioresistant Bacterium Deinococcus
RT radiodurans R1."
RL Science 286:1571-1577(1999).
DR EMBL: AE002079; AAF12045.1; -.
DR TIGR: DR2500; -.
KW Hypothetical protein.
SQ SEQUENCE 157 AA; 17027 MW; B766BD89F60A5B5D CRC64;

Query Match 33.9%; Score 41; DB 2; Length 157;
Best Local Similarity 42.1%; Pred. No. 8.1;
Matches 8; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 16 EEAVRLXXXLKNKGXSSG 34
DB 74 DDAVOVFYRLAKNAGLDSG 92

RESULT 5
OQ2143 PRELIMINARY; PRT; 266 AA.
AC OQ2143;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE PROGLUCAGON I.
OS Xenopus laevis (African clawed frog).

```



OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
 OC Xenopodidae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97368292; PubMed=9223287;  
 RA Irwin D.M., Satkunaratjah M., Wen Y., Brubaker P.L., Pederson R.A.,  
 Wheeler M.B.;  
 RT "The xenopus progulacagon gene encodes novel GLP-1-like peptides with  
 insulinotropic properties.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 94:7915-7920(1997).  
 DR EMBL; AF004432; AACB5660.1; -.  
 DR HSSP; P01274; 1GCM.  
 DR INTERPRO; IPR000532; -.  
 DR PFAM; PF00123; hormone2; 5.  
 DR PRINTS; PR00275; GLUCAGON.  
 DR PROSITE; PS00260; GLUCAGON; 5.  
 DR PRODOM; PD002392; -. 1.  
 DR PRODOM; PD002700; -. 3.  
 SQ SEQUENCE 266 AA; 30951 MW; 544F7B8C20AF872C CRC64;

Query Match 33.9%; Score 41; DB 13; Length 266;  
 Best Local Similarity 34.5%; Pred. No. 15;  
 Matches 10; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

OY 4 GTXXXXXSKOXEEAVRLXXXXKNGXGS 32  
 DB 100 GTTSDVTQQLDERAKAFIDWLINGPS 128

RESULT 6  
 ID 092527 PRELIMINARY; PRT; 306 AA.

AC 092527;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE COAT PROTEIN (CAPSID PROTEIN).  
 OS Carnation latent virus (CLV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.  
 OX NCBI\_TaxID=12164;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Meehan B.M.;  
 RL Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91324119; PubMed=1713905;  
 RA Meehan B.M., Mills P.R.;  
 RT "Nucleotide sequence of the 3'-terminal region of carnation latent  
 virus.";  
 RL Intervirology 32:262-267(1991).  
 DR EMBL; AJ010697; CA009306.1; -.  
 DR INTERPRO; IPR000052; -.  
 DR PFAM; PF00286; virus\_P-coat; 1.  
 DR PRINTS; PR00232; POTXCARLCOAT.  
 DR PROSITE; PS00418; POTEX\_CARLAVIRUS\_COAT; 1.  
 DR PRODOM; PD000603; -. 1.  
 DR Coat protein.  
 SQ SEQUENCE 306 AA; 33890 MW; 4456BB53E174298 CRC64;

Query Match 33.9%; Score 41; DB 12; Length 306;  
 Best Local Similarity 43.5%; Pred. No. 17;  
 Matches 10; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

OY 12 KOXEEAVRLXXXXKNGXSSG 34  
 DB 62 KLEKERNSONLTAGELKNGGFESG 84

RESULT 7  
 ID 09UYT6 PRELIMINARY; PRT; 402 AA.  
 AC 09UYT6;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE MOLYBDENUM COFACTOR BIOSYNTHESIS PROTEIN (MOEA-1).  
 GN PAB1436.  
 OS Pyrococcus abyssi.  
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
 OX NCBI\_TaxID=29292;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ORSAV;  
 RA Hellig R.;  
 RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome  
 structure and evolution.";  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ248287; CAB50326.1; -.  
 DR INTERPRO; IPR001453; -.  
 DR INTERPRO; IPR002106; -.  
 DR PFAM; PF00994; MOCF\_biosynth; 1.  
 DR PROSITE; PS00339; AA\_TRNA\_LIGASE\_II\_2; UNKNOWN\_1.  
 SQ SEQUENCE 402 AA; 43327 MW; 44545EDA70F6A78E CRC64;

Query Match 33.9%; Score 41; DB 1; Length 402;  
 Best Local Similarity 39.1%; Pred. No. 23;  
 Matches 9; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

OY 12 KOXEEAVRLXXXXKNGXSSG 34  
 DB 237 KLEIEGVRVADIVVISGSASGG 259

RESULT 8  
 ID 075464 PRELIMINARY; PRT; 589 AA.

AC 075464;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE NUCLEAR MATRIX PROTEIN NRP/B.  
 GN NRPB.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98234394; PubMed=9566959;  
 RA Kim T.A., Lim J., Ota S., Raja S., Rogers R., Rivnay B., Avraham H.,  
 Avraham S.;  
 RT "NRP/B, a novel nuclear matrix protein, associates with p110(Rb) and  
 is involved in neuronal differentiation.";  
 RL J. Cell Biol. 141:553-566(1998).  
 DR EMBL; AF059611; AAC26109.1; -.  
 DR INTERPRO; IPR000169; -.  
 DR INTERPRO; IPR000210; -.  
 DR INTERPRO; IPR001798; -.  
 DR PFAM; PF00651; BFB; 1.  
 DR PFAM; PF01344; Kelch; 6.  
 DR PROSITE; PS00639; THIOL\_PROTEASE\_HIS; UNKNOWN\_1.  
 DR Matrix protein.  
 SQ SEQUENCE 589 AA; 66129 MW; DE003A1DFA65BA0 CRC64;

Query Match 33.9%; Score 41; DB 4; Length 589;  
 Best Local Similarity 45.0%; Pred. No. 36;  
 Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

OY 11 SKOXEEAVRLXXXXKNGG 30

Db 262 SKEIVEAIRCKIKLONDG 281

```

RESULT 9
ID Q9UPG9 PRELIMINARY: PRT: 589 AA.
AC Q9UPG9;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE ECTODERM-NEURAL CORTEX-1 PROTEIN.
GN ENC-1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98350113; PubMed=9683534;
RA Hernandez M.-C., Andres-Barquin P.J., Holt I., Israel M.A.;
RT "Cloning of human ENC-1 and evaluation of its expression and
RT regulation in nervous system tumors.";
RL Exp. Cell Res. 242:470-477(1998).
DR EMBL: AF005381; AAC64498.1; -.
DR INTERPRO: IPR000169; -.
DR INTERPRO: IPR000210; -.
DR INTERPRO: IPR001798; -.
DR PFAM: PF00651; BTB: 1.
DR PFAM: PF01344; Kelch: 6.
DR PROSITE: PS00639; THIOLEPROTEASE_HIS; UNKNOWN_1.
SQ SEQUENCE 589 AA; 6613 MW; C0002116EA6A1AB CRC64;

```

Query Match 33.9%; Score 41; DB 4; Length 589;  
 Best Local Similarity 45.0%; Pred. No. 36;  
 Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

```

OY 11 SKOXEEAVRLXXXXLKNKG 30
DB 262 SKEIVEAIRCKIKLONDG 281

RESULT 10
ID Q35709 PRELIMINARY: PRT: 589 AA.
AC Q35709;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE ECTODERM-NEURAL CORTEX-1 PROTEIN (ENC-1).
GN ENCL OR ENC-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ALBINO SWISS; TISSUE=BRAIN;
RA MEDLINE=97252647; PubMed=9096139;
RA Hernandez M.-C., Andres-Barquin P.J., Martinez S., Bulfone A.,
RA Rubenstein J.L.R., Israel M.A.;
RA "ENC-1: a novel mammalian kelch-related gene specifically expressed in
RT the nervous system encodes an actin-binding protein.";
RL J. Neurosci. 17:3038-3051(1997).
CC - FUNCTION: ACTIN-BINDING PROTEIN INVOLVED IN THE REGULATION OF
CC NEURONAL PROCESS FORMATION AND IN DIFFERENTIATION OF NEURAL CREST
CC CELLS.
CC - TISSUE SPECIFICITY: PRIMARILY EXPRESSED IN THE NERVOUS SYSTEM.
CC - DEVELOPMENTAL STAGE: EXPRESSION IS HIGHLY DYNAMIC BUT MOSTLY
CC RESTRICTED TO THE NS. OUTSIDE THE NS, EXPRESSION IS DETECTED IN
CC THE ROSTRAL-MOST SOMITOMERE OF THE PRESONITIC MESODERM, AT THE
CC TIMES CORRESPONDING TO THE EPITHELIALIZATION THAT PRECEDES SOMITE

```

```

CC CC FORMATION. FIRST DETECTED IN THE BRAIN AND SPINAL CHORD OF 12 PC
CC CC EMBRYOS.
CC - SUBCELLULAR LOCATION: CYTOPLASMIC. INTERACTS WITH THE ACTIN
CC CYOSKELETON.
CC - SIMILARITY: BELONGS TO THE KELCH/MLP FAMILY.
CC - SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
DR EMBL: U63079; AAB64206.1; -.
DR MGD: MGI:109610; Encl.
DR INTERPRO: IPR000210; -.
DR INTERPRO: IPR001798; -.
DR PFAM: PF00651; BTB: 1.
DR PFAM: PF01344; Kelch: 6.
DR Phosphorylation; Actin-binding; Developmental protein; Cytoskeleton.
KW DOMAIN
FT 28 144
FT 296 585
FT REPEAT 296 340 1.
FT REPEAT 341 388 2.
FT REPEAT 389 444 3.
FT REPEAT 445 492 4.
FT REPEAT 493 538 5.
FT REPEAT 539 585 6.
SQ SEQUENCE 589 AA; 66085 MW; 12E62354D508B6A2 CRC64;

```

Query Match 33.9%; Score 41; DB 11; Length 589;  
 Best Local Similarity 45.0%; Pred. No. 36;  
 Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

OY 11 SKOXEEAVRLXXXXLKNKG 30  
 DB 262 SKEIVEAIRCKIKLONDG 281

```

RESULT 11
ID Q9VRN7 PRELIMINARY: PRT: 2630 AA.
AC Q9VRN7;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE CG5256 PROTEIN.
GN CG5256.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Adamatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Landell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew K.Y., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borovaya D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Butlis K.C., Busam D.A., Butler H., Cadieu L.B., Davies P.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glöckl A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.D., Harvey D., Heilmann T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodish C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

```

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mlshina N.V., Mobarri C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Murry D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003565; AAF50755.1; -.  
 DR HSSP: P08567; 1PLS.  
 DR FLYBASE: FBgn0035628; CG5256.  
 DR INTERPRO: IPR000219; -.  
 DR INTERPRO: IPR001331; -.  
 DR INTERPRO: IPR001478; -.  
 DR INTERPRO: IPR001849; -.  
 DR INTERPRO: IPR003116; -.  
 DR PFAM: PF00169; PH: 2.  
 DR PFAM: PF00621; RhogEF; 1.  
 DR PFAM: PF02196; RBD; 1.  
 DR PROSITE: PS00741; GDS\_CDC24; UNKNOWN\_1.  
 DR PROSITE: PS50003; PH\_DOMAIN; 1.  
 SQ SEQUENCE 2630 AA; 292511 MW; 56D1AD5971FE5B9 CRC64;

Query Match 33.9%; Score 41; DB 5; Length 2630;  
 Best Local Similarity 41.7%; Pred. No. 1.9e+02;  
 Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

OY 12 KQXEENAVRLXXXXKNGXSSGA 35  
 DB 1762 RQIRRESVRNMSIPMKNGSSGS 1785

RESULT 12  
 ID 09VRN8 PRELIMINARY; PRT; 2637 AA.  
 AC 09VRN8;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE CG5406 PROTEIN.  
 GN CG5406.  
 OS *Drosophila melanogaster* (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_Taxid=1227;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Chame M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agayuni A., An H.-J., Andrews-Plamkoc C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu T., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Gang N.S., Gelbart W.M., Glasser K.,  
 RA Glosier A., Gorgi F., Gorrell J.H., Gu Z., Guan P., Harris K.,  
 RA Harris N.L., Harey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodali C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mlshina N.V., Mobarri C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Murry D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003565; AAF50755.1; -.  
 DR HSSP: P08567; 1PLS.  
 DR FLYBASE: FBgn0035629; CG5406.  
 DR INTERPRO: IPR000219; -.  
 DR INTERPRO: IPR001331; -.  
 DR INTERPRO: IPR001478; -.  
 DR INTERPRO: IPR001849; -.  
 DR INTERPRO: IPR003116; -.  
 DR PFAM: PF00169; PH: 2.  
 DR PFAM: PF00621; RhogEF; 1.  
 DR PFAM: PF02196; RBD; 1.  
 DR PROSITE: PS00741; GDS\_CDC24; UNKNOWN\_1.  
 DR PROSITE: PS50003; PH\_DOMAIN; 1.  
 SQ SEQUENCE 2637 AA; 293079 MW; 76F475BA3CBE961B CRC64;

Query Match 33.9%; Score 41; DB 5; Length 2637;  
 Best Local Similarity 41.7%; Pred. No. 1.9e+02;  
 Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

OY 12 KQXEENAVRLXXXXKNGXSSGA 35  
 DB 1769 RQIRRESVRNMSIPMKNGSSGS 1792

RESULT 13  
 ID P96631 PRELIMINARY; PRT; 127 AA.  
 AC P96631;  
 DT 01-MAY-1997 (TREMBLrel. 03, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE PROBABLE REPRESSOR PROTEIN.  
 GN YDCN.  
 OS *Bacillus subtilis*.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_Taxid=1423;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98000887; PubMed=9341680;  
 RA Beloin C., Ayora S., Exley R., Hirschbein L., Ogasawara N.,  
 RA Kasahara Y., Alonso J.C., Le Hegarat F.;  
 RT "Characterization of an *lfp*-like (*lfpC*) gene from *Bacillus subtilis*.";  
 RL Mol. Gen. Genet. 256:63-71(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;

```

RX MEDLINE:98044033; PubMed:9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Avevedo V., Bertero M.G., Bessieres P., Bolotin A., Borcherdt S.,
RA Borris R., Bourcier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel K.A.,
RA Denicot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrati E., Galleron N.,
RA Fritz C.Y., Fujita M., Fujita Y., Fuma S., Galiziz A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grand G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henat A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaere-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazerec V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Mostel D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Portwolk S., Prescott A.M.,
RA Pressac E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scottone F.,
RA Sekiuchi J., Sekowska A., Serot S.J., Serron P., Shin B.S., Soldo B.,
RA Sorokin A., Taconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Vtari A., Wambolt R., Wedler E., Wedler H., Weitzengerger T.,
RA Winters P., Wipert A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (Nov-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB001488; BAA19320.1; -
DR EMBL: Z99106; CAB12289.1; -
DR INTERPRO: IPR001387; -
DR PRAM: PFI01381; HTB_3; 1.
SQ SEQUENCE 127 AA; 14649 MW; 3CC91D5B1D51628C CRC64;

Query Match 33.1%; Score 40; DB 2; Length 127;
Best Local Similarity 47.1%; Pred. No. 9.7;
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 13 QXEEAVRLXXXXLKG 29
Db 100 EFDETFARLVKKALKG 116

RESULT 14
O9U184 PRELIMINARY; PRT; 374 AA.
AC O9U184;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE HYPOHETICAL 41.7 KDA PROTEIN.
GN L3238.06.
OS Leishmania major.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FRIEDLIN;
RA Medler H., Hilbert H., Duesterhoft A., Ivens A.C., Murphy L.,
RA Quail M., Rajandream M.A., Barrell B.G.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-FRIEDLIN;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome."
RL Genome Res. 8:135-145(1998).
DR EMBL: AL133468; CAB63133.1; -
RW Hypothetical protein.
SQ SEQUENCE 374 AA; 41675 MW; AA38847DBE433937 CRC64;

Query Match 33.1%; Score 40; DB 5; Length 374;
Best Local Similarity 39.1%; Pred. No. 33;
Matches 9; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

OY 12 KQXEEAVRLXXXXLKGXSSG 34
Db 14 KKKKKKKKASOGPLRGAGSSG 36

RESULT 15
O9LHL3 PRELIMINARY; PRT; 455 AA.
AC O9LHL3;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE RNA-BINDING PROTEIN-LIKE.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (May-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety pl,
RT TAC and BAC clones."
RL DNA Res. 7:217-221(2000).
DR EMBL: AF002041; BAB02607.1; -
SQ SEQUENCE 455 AA; 48958 MW; E88117B2C33BBA9 CRC64;

Query Match 33.1%; Score 40; DB 10; Length 455;
Best Local Similarity 41.7%; Pred. No. 41;
Matches 10; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

OY 12 KQXEEAVRLXXXXLKGXSSG 35
Db 5 QREVEVSYTEKKRLDGGSSG 28

Search completed: February 13, 2001, 16:07:25
Job time: 80 sec

```



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2001, 16:06:44 ; Search time 12.65 Seconds

(without alignments)  
102.116 Million cell updates/sec

Title: US-08-908-867-38

Perfect score: 121  
Sequence: 1 XXXGTXXXXXXSRQXEEAVRLXXXXLKNKGXSSGAXXXXX 40

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	91	75.2	39	1	EXE3_HELHO
2	91	75.2	87	1	EXE4_HELSD
3	42	34.7	546	1	PGMU_ECOLI
4	41	33.9	516	1	P110_HUMAN
5	41	33.9	2044	1	SIF2_DROME
6	41	33.9	2064	1	SIF1_DROME
7	39	32.2	341	1	HRP2_PSEST
8	38	31.4	300	1	TF2B_PYRHO
9	38	31.4	355	1	NDF1_MESAU
10	38	31.4	357	1	NDF1_MOUSE
11	38	31.4	357	1	NDF1_RAT
12	38	31.4	419	1	DNLI_ASPM2
13	38	31.4	589	1	LOIA_LYCPN
14	37	30.6	430	1	AST2_YEAST
15	37	30.6	488	1	KPT1_CAEBL
16	37	30.6	1237	1	KPB1_RABIT
17	36	29.8	318	1	NSR_LACLA
18	36	29.8	324	1	GLXA_RHIME
19	36	29.8	324	1	VP35_VACCC
20	36	29.8	325	1	VP35_VARY
21	36	29.8	373	1	BIOE_AOUAE
22	36	29.8	401	1	CPXP_BRAJA
23	36	29.8	413	1	FLI_TOBAC
24	36	29.8	633	1	SOHR_RHIME
25	36	29.8	845	1	SCPI_MESAU
26	36	29.8	3068	1	POLG_PEMVC
27	36	29.8	4687	1	PLEC_RAT
28	35.5	29.3	111	1	HMGZ_DROME
29	35	28.9	261	1	TF2B_PYRMO
30	35	28.9	273	1	FGC_STRMU
31	35	28.9	300	1	TF2B_PYRPU
32	35	28.9	356	1	NDF1_HUMAN
33	35	28.9	416	1	FL2_TOBAC

34	35	28.9	460	1	VP41_BPAPS
35	35	28.9	467	1	HSLO_BACSU
36	35	28.9	501	1	FLAA_AOUPI
37	35	28.9	509	1	GATB_MYCLE
38	35	28.9	563	1	IDS_MOUSE
39	35	28.9	639	1	CALC_RABIT
40	35	28.9	748	1	HEPA_HSYBC
41	35	28.9	759	1	NAHB_ONCMY
42	35	28.9	858	1	H105_CRICR
43	35	28.9	938	1	NMZ1_HUMAN
44	35	28.9	938	1	NMZ1_MOUSE
45	35	28.9	938	1	NMZ1_RAT

## ALIGNMENTS

RESULT 1					
EXE3_HELHO	STANDARD:	PRT:	39 AA.		
AC P20394:					
DT 01-FEB-1991 (Rel. 17, Created)					
DT 01-FEB-1991 (Rel. 17, Last sequence update)					
DT 01-MAY-1992 (Rel. 22, Last annotation update)					
DE EXENDIN-3.					
OS Heloderma horridum horridum (Mexican beaded lizard).					
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC Lepidosauria; Squamata; Scleroglossa; Anguilliformes; Helodermatidae;					
OC Heloderma.					
RN [1]					
RP SEQUENCE.					
RC TISSUE=VENOM;					
RX MEDLINE=91056067; PubMed=1700785;					
RA Eng J., Andrew P.C., Kleiman W.A., Singh L., Rautman J.-P.					
RT "Purification and structure of exendin-3, a new pancreatic					
secretagogue isolated from Heloderma horridum venom."					
RL J. Biol. Chem. 265:20259-20262(1990).					
CC -1- FUNCTION: HAS A VIP/SECRETIN-LIKE BIOLOGICAL ACTIVITY. INTERACTS					
CC WITH THE EXENDIN RECEPTOR.					
CC -1- SIMILARITY: BELONGS TO THE GLUCAGON FAMILY.					
DR PIR: A23674; HMG32.					
DR HSSP; P01274; IGCN.					
DR INTERPRO: IPR000532; -1					
DR PFAM: PF00123; hormone2; 1.					
DR PROSITE: PS00260; GLUCAGON; 1.					
DR Glucagon family; Venom; Amidation.					
FT MOD.RES 39					
FT SEQUENCE 39 AA; 4204 MW; A44251D5A4B1D1B9 CRC64;					
Query Match	75.2%;	Score 91;	DB 1;	Length 39;	
Best local similarity	65.6%;	Pred. No. 4.7e-10;			
Matches 21; Conservative 0;	Mismatches 11;	Indels 0;	Gaps 0;		
OY 4 GTXXXXXSRQXEEAVRLXXXXLKNKGXSSGA 35					
DB 4 GFTSDLSKQMEAEVRLFIETLKNKGSSGA 35					
RESULT 2					
EXE4_HELSD	STANDARD:	PRT:	87 AA.		
AC P26349:					
DT 01-MAY-1992 (Rel. 22, Created)					
DT 15-JUL-1998 (Rel. 36, Last sequence update)					
DT 30-MAY-2000 (Rel. 39, Last annotation update)					
DE EXENDIN-4 PRECURSOR.					
OS Heloderma suspectum (Gila monster).					
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC Lepidosauria; Squamata; Scleroglossa; Anguilliformes; Helodermatidae;					
OC Heloderma.					
RN [1]					
RP SEQUENCE FROM N.A.					

```

RX MEDLINE-97172477; PubMed-9020121;
RA Chen Y.E., Drucker D.J.;
RT "Tissue-specific expression of unique mRNAs that encode proglucagon-
RT derived peptides or exendin 4 in the lizard.";
RL J. Biol. Chem. 272:4108-4115(1997).
RN [2]
RP SEQUENCE OF 48-86.
RC TISSUE-VENOM;
RX MEDLINE-92218391; PubMed-1313797;
RA Eng J., Kleiman W.A., Singh L., Singh G., Raufman J.-P.;
RT "Isolation and characterization of exendin-4, an exendin-3 analogue,
RT from Heloderma suspectum venom. Further evidence for an exendin
RT receptor on dispersed acini from guinea pig pancreas.";
RL J. Biol. Chem. 267:7402-7405(1992).
CC -1- FUNCTION: HAS A VIP/SECRETIN-LIKE BIOLOGICAL ACTIVITY. INTERACTS
CC WITH THE EXENDIN RECEPTOR.
CC -1- SIMILARITY: BELONGS TO THE GLUCAGON FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U77613; AAB51130.1; -.
DR PIR: A42486; HMHGAG.
DR INTERPRO: IPR000532; -.
DR PFM: PFO0123; hormone2; 1.
DR PROSITE: PS00260; GLUCAGON; 1.
KW Glucagon family; Venom; Amidation; Signal.
FT SIGNAL 1 23
FT PEPTIDE 48 86 EXENDIN-4.
FT MOD_RES 86 86 AMIDATION (G-87 PROVIDE AMIDE GROUP).
SQ SEQUENCE 87 AA: 9479 MW: 656BA6E3D87454A2 CRC64;

```

```

Query Match 75.2%; Score 91; DB 1; Length 87;
Best Local Similarity 65.6%; Pred. No. 1; le-09;
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

```

```

OY 4 GTXXXXXKQXEEAVRLXXXXKNGXSSGA 35
DB 51 GFTSDLSKQXEEAVRLFEIWLKNGSPSSGA 82

```

RESULT 3

PGMU\_ECOLI STANDARD; PRT; 546 AA.

AC P36938;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE PHOSPHOGLUCOMUTASE (EC 5.4.2.2) (GLUCOSE PHOSPHOMUTASE) (PGM).

GN PGM.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-K12;

RX MEDLINE-94364967; PubMed-8083177;

RA Lu M., Kleckner N.;

RT "Molecular cloning and characterization of the pgm gene encoding

RT phosphoglucumutase of Escherichia coli.";

RL J. Bacteriol. 176:5847-5851(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-K12 / MG1655;

RX MEDLINE-97426617; PubMed-9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

```

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE-97061202; PubMed-8905232;
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizouchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sangei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN [4]
RP SEQUENCE OF 1-20 FROM N.A.
RC STRAIN-K12;
RX MEDLINE-94236686; PubMed-8011018;
RA Lu M., Campbell J.L., Boye E., Kleckner N.;
RT "SegA: a negative modulator of replication initiation in E. coli.";
RL Cell 77:413-426(1994).
RN [5]
RP CHARACTERIZATION.
RA Joshi J.G., Handler P.;
RT "Phosphoglucumutase. II. Purification and properties of
RT phosphoglucumutase from Escherichia coli.";
RL J. Biol. Chem. 239:2741-2751(1964).
CC -1- FUNCTION: THIS ENZYME PARTICIPATES IN BOTH THE BREAKDOWN AND
CC SYNTHESIS OF GLUCOSE.
CC -1- CATALYTIC ACTIVITY: ALPHA-D-GLUCOSE 1-PHOSPHATE = ALPHA-D-GLUCOSE
CC 6-PHOSPHATE.
CC -1- SIMILARITY: TO OTHER PHOSPHOGLUCOMUTASES AND PHOSPHOMANNOMUTASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U08369; AAA57067.1; -.
DR EMBL: AE000172; AAC73782.1; -.
DR EMBL: D90707; BAA35337.1; -.
DR EMBL: D90708; BAA35345.1; -.
DR EMBL: U07651; -. NOT_ANNOTATED_CDS.
DR ECOGENE: EG12144; PGM.
DR INTERPRO: IPR001485; -.
DR PFM: PFO0408; PGM_PGM; 1.
DR PROSITE: PS00710; PGM_PGM; 1.
KW Isomerase; Phosphorylation.
FT ACT_SITE 146 146
FT SEQUENCE 546 AA: 58361 MW: 666B6B9C2FE2ECD59 CRC64;

```

```

Query Match 34.7%; Score 42; DB 1; Length 546;
Best Local Similarity 52.9%; Pred. No. 4; 8;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

```

```

OY 12 KQXEEAVRLXXXXLKN 28
DB 529 KQIEKAVIYSEVLKN 545

```

RESULT 4

P110\_HUMAN STANDARD; PRT; 516 AA.

ID P110\_HUMAN

AC O14682;

DT 30-MAY-2000 (Rel. 39, Created)



DT 30-MAY-2000 (Rel. 39, last sequence update)  
 DT 30-MAY-2000 (Rel. 39, last annotation update)  
 RA P53-INDUCED PROTEIN 10.  
 DE P53-INDUCED PROTEIN 10.  
 GN P53.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 CC [1]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE-COLON CANCER;  
 RX MEDLINE-97449378; PubMed-9305847;  
 RA Polyak K., Xia Y., Zweier J.L., Kinzler K.W., Vogelstein B.;  
 RT Nature 389:300-306(1997).  
 CC -1- SIMILARITY: BELONGS TO THE KILCH/MIP FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC  
 CC EMBL; AF010314; AAC39532.1; -  
 DR INTERPRO: IPR000169; -  
 DR INTERPRO: IPR000210; -  
 DR INTERPRO: IPR001798; -  
 DR PFM; PF00651; BTB; 1.  
 DR PFM; PF01344; Kelch; 3.  
 DR PROSITE: PS50097; BTB; 1.  
 KW Repeat.  
 FT DOMAIN 30 147 BTB.  
 FT REPEAT 299 447 3 APPROXIMATE KELCH TANDEM REPEATS.  
 FT REPEAT 349 343 1.  
 FT REPEAT 391 2.  
 FT REPEAT 392 447 3.  
 SO SEQUENCE 516 AA; 58314 MW; DFC3D773B3D3B9B2 CRC64;

Query Match 33.9%; Score 41; DB 1; Length 516;  
 Best Local Similarity 45.0%; Pred. No. 6.9;  
 Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 11 SKXEEAVRLKXXKNG 30  
 DB 265 SKIYEAIKCKIKLQNDG 284  
 ID SIF2\_DROME STANDARD: PRT; 2044 AA.  
 AC P91620;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, last sequence update)  
 DT 15-JUL-1999 (Rel. 38, last annotation update)  
 DE STILL LIFE PROTEIN TYPE 2 (SIF TYPE 2).  
 GN SIF.  
 OS Drosophila melanogaster (Fruit fly).  
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC Ephydroidea; Drosophilidae; Drosophila.  
 CC [1]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE-HEAD;  
 RX MEDLINE-97153054; PubMed-8999801;  
 RA Sone M., Hoshino M., Suzuki E., Kuroda S., Kaibuchi K., Nakagoshi H.,  
 RA Saigo K., Nabeshima Y.-I., Hama C.;  
 RT "Still life, a protein in synaptic terminals of Drosophila homologous  
 RT to GDP-GTP exchangers";  
 RL Science 275:543-547(1997).  
 RN [2]

RP ERRATUM.  
 RA Sone M., Hoshino M., Suzuki E., Kuroda S., Kaibuchi K., Nakagoshi H.,  
 RA Saigo K., Nabeshima Y.-I., Hama C.;  
 RL Science 275:1405-1405(1997).  
 CC -1- FUNCTION: REGULATES SYNAPTIC DIFFERENTIATION THROUGH THE  
 CC ORGANIZATION OF ACTIN CYTOSKELETON POSSIBLY BY ACTIVATING RHO-LIKE  
 CC GTPASES. IS LIKELY A FACTOR IN THE CASCADE OF RAC1 OR CDC42 IN THE  
 CC NEURONS.  
 CC -1- SUBCELLULAR LOCATION: LOCALIZES TO THE SUBMEMBRANOUS REGION OF  
 CC SYNAPTIC TERMINALS.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS, SIF TYPE 1 (P91621) AND SIF TYPE  
 CC 2 (SHOWN HERE), ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- DEVELOPMENTAL STAGE: AT STAGE 14, EXPRESSION OCCURS IN EACH  
 CC SEGMENT OF THE CENTRAL NERVOUS SYSTEM. AT STAGE 17, EXPRESSION  
 CC BECOMES RESTRICTED TO THE SYNAPTIC REGIONS OF THE BRAIN AND  
 CC VENTRAL NERVE CORD, WHERE SYNAPSES UNDERGO MATURATION.  
 CC  
 CC -1- SIMILARITY: CONTAINS 1 DEL-HOMOLOG DOMAIN (DH).  
 CC  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC  
 CC EMBL; D86546; BAA13108.1; -  
 DR FLXBASE; FBgn0018652; sif.  
 DR INTERPRO: IPR000219; -  
 DR INTERPRO: IPR001331; -  
 DR INTERPRO: IPR001478; -  
 DR INTERPRO: IPR001849; -  
 DR PFM; PF00595; PDZ; 1.  
 DR PFM; PF00169; PH; 2.  
 DR PFM; PF00621; Rhogef; 1.  
 DR PROSITE: PS00741; GDS\_CDC24; 1.  
 DR PROSITE: PS50003; PH\_DOMAIN; 1.  
 KW Guanine-nucleotide releasing factor; Developmental protein; Synapse;  
 KW Repeat; Alternative splicing.  
 FT DOMAIN 62 249 4 X 25 AA APPROXIMATE REPEAT.  
 FT REPEAT 62 86 1.  
 FT REPEAT 94 118 2.  
 FT REPEAT 154 178 3.  
 FT REPEAT 225 249 4.  
 FT DOMAIN 819 937 PH.  
 FT DOMAIN 1184 1273 PDZ.  
 FT DOMAIN 1410 1673 DH.  
 FT DOMAIN 1674 1767 PH.  
 FT DOMAIN 467 470 POLY-PRO.  
 FT DOMAIN 646 649 POLY-ARG.  
 FT DOMAIN 1295 1298 POLY-PRO.  
 FT DOMAIN 1898 1909 POLY-GLN.  
 FT DOMAIN 1929 1933 POLY-PRO.  
 SO SEQUENCE 2044 AA; 228324 MW; 75D7CF21F49654B6 CRC64;

Query Match 33.9%; Score 41; DB 1; Length 2044;  
 Best Local Similarity 41.7%; Pred. No. 29;  
 Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 12 KOXEEAVRLKXXKNGSSGA 35  
 DB 1761 RQIRSVNMSIPKRNFGSSGS 1784  
 ID SIF1\_DROME STANDARD: PRT; 2064 AA.  
 AC P91621;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, last sequence update)  
 DT 15-OCT-2000 (Rel. 40, last annotation update)



[1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-013;  
 RX MEDLINE-96344137; PubMed-9679194;  
 RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,  
 RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohnuku Y.,  
 RA Funahashi T., Tanaka T., Kudo Y., Yamazaki J., Kishida N., Oguchi A.,  
 RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,  
 RA Mauchl Y., Shizuya H., Kikuchi H.;  
 RT "Complete sequence and gene organization of the genome of a hyper-  
 thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";  
 RL DNA Res. 5:55-76(1998).  
 CC -1- FUNCTION: STABILIZES TBP BINDING TO AN ARCHAEAL BOX-A PROMOTER.  
 CC ALSO RESPONSIBLE FOR RECRUITING RNA POLYMERASE II TO THE PRE-  
 CC INITIATION COMPLEX (DNA-TBP-TFIIB) (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE TFIIB FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AP000006; BAA30589.1; -.  
 DR INTERPRO: IPR000812; -.  
 DR PFAM: PF00382; transcript\_fac2; 2.  
 DR PRINTS: PRO0685; TIFACTOR1IB.  
 DR PROSITE: PS00782; TFIIB; 2.  
 KW Transcription regulation; Repeat; Zinc-finger.  
 FT ZN.FING  
 FT 7  
 SQ SEQUENCE 300 AA; 34097 MW; DE3758F398BC855F CRC64;

Query Match 31.4%; Score 38; DB 1; Length 300;  
 Best Local Similarity 44.4%; Pred. No. 14;  
 Matches 8; Conservative 2; Mismatches 8; Indels 0; Gaps 0;  
 QY 12 KXEEAVRLXXXXKNG 29  
 DB 127 KXEEAVRLXXXXKNG 144  
 ID NDF1\_MOUSE STANDARD; PRT; 355 AA.  
 AC Q60430;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE NEUROGENIC DIFFERENTIATION FACTOR 1 (BETA-CELL E-BOX TRANS-ACTIVATOR  
 DE 2) (BETA2).  
 GN NEUROD1 OR NEUROD.  
 OS Mesocricetus auratus (Golden hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Mesocricetus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-95293222; PubMed-7774807;  
 RA Naya F.J., Stellrecht C.M.M., Tsai M.-J.;  
 RT "Tissue-specific regulation of the insulin gene by a novel basic  
 RT helix-loop-helix transcription factor.";  
 RL Genes Dev. 9:1009-1019(1995).  
 CC -1- FUNCTION: ACTS AS A DIFFERENTIATION FACTOR DURING NEUROGENESIS.  
 CC TRANSCRIPTIONAL ACTIVATOR. BINDS TO THE INSULIN GENE E-BOX.  
 CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN. HETERODIMER WITH E47.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
 CC -1- TISSUE SPECIFICITY: MOST ABUNDANT IN PANCREATIC ALPHA- AND BETA-  
 CC CELLS, LESS IN BRAIN AND INTESTINE.

-1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS. "ATONAL" SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U24679; AAA86518.1; AL1-INT.  
 DR HSSP: P10085; IMDY.  
 DR INTERPRO: IPR001092; -.  
 DR INTERPRO: IPR003015; -.  
 DR PFAM: PF00010; HDH; 1.  
 DR PROSITE: PS00038; HELIX\_LOOP\_HELIX; 1.  
 KW DNA-binding; Nuclear protein; Transcription regulation; Activator;  
 KW Neurogenesis; Developmental protein; Differentiation.  
 FT DOMAIN 58  
 FT 86 92 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT 101 112 BASIC DOMAIN.  
 FT DOMAIN 113 153 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
 FT 67 75 POLY-GLU.  
 FT DOMAIN 86 89 POLY-LYS.  
 SQ SEQUENCE 355 AA; 39763 MW; FA344DFD360226B2 CRC64;

Query Match 31.4%; Score 38; DB 1; Length 355;  
 Best Local Similarity 42.1%; Pred. No. 16;  
 Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;  
 QY 12 KXEEAVRLXXXXKNG 30  
 DB 39 KXEEAVRLXXXXKNG 57  
 ID NDF1\_MOUSE STANDARD; PRT; 357 AA.  
 AC Q60867; Q60897;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE NEUROGENIC DIFFERENTIATION FACTOR 1.  
 GN NEUROD1 OR NEUROD.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN-MEL AND 129/SV;  
 RX MEDLINE-95273957; PubMed-7754368;  
 RA Lee J.E., Hollenberg S.M., Snider L., Turner D.L., Lipnick N.,  
 RA Weintraub H.;  
 RT "Conversion of Xenopus ectoderm into neurons by NeuroD, a basic  
 RT helix-loop-helix protein.";  
 RL Science 266:836-844(1995).  
 CC -1- FUNCTION: ACTS AS A DIFFERENTIATION FACTOR DURING NEUROGENESIS.  
 CC TRANSCRIPTIONAL ACTIVATOR. BINDS TO THE INSULIN GENE E-BOX.  
 CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN. HETERODIMER WITH E47.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN DIFFERENTIATING NEURONS OF  
 CC BOTH THE CENTRAL AND PERIPHERAL NERVOUS SYSTEMS.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING EMBRYONIC DEVELOPMENT.  
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS. "ATONAL" SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way



Query Match	31.48;	Score 38;	DB 1;	Length 589;
Best Local Similarity	36.08;	Pred. No. 27;		

RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.  
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,

RA Fullon L, Gardner A., Green P., Hawkins T., Haller L., Jier M.,  
 RA Johnston L., Jones M., Kershaw J., Kirsten J., Rittiser N.,  
 RA Larrelle P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifken L., Koopra A., Saunders E., Showkheen R.,  
 RA Sims M., Smaldon N., Smith A., Smith M., Sonnenhammer E., Staden R.,  
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,  
 RA Waterson R., Watson A., Weinstock L., Wilkinson-Spratt J.,  
 RA Woldman P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RL elegans." Nature 368:32-38(1994).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement>).  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z29094; GAB82340.1; -.  
 DR PIR: S40706; S40706.  
 DR HSSP: P23807; 11XX.  
 DR WORMPEP: C07A9.1; CE00502.  
 DR INTERPRO: IPR001304; -.  
 DR PFAM: PF00059; lectin\_c.1.  
 DR PROSITE: PS00041; C\_Type\_Lectin\_2; UNKNOWN\_1.  
 DR Hypothetical protein.  
 SO Sequence 488 aa; 54717 MW; 9022651E47078814 CRC64;

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2001, 16:06:05 ; Search time 21.65 Seconds

(without alignments)  
125.452 Million cell updates/sec

Title: US-08-908-867-38

Perfect score: 121  
Sequence: 1 XXXGTXXXXXXSKQXEEAVRLXXXXLKNKGXSSGAXXXXX 40

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_66:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	91	75.2	39	1 HMGH3Z	exendin-3 - Mexica
2	91	75.2	39	1 HMGH3Z	exendin-4 - Gila m
3	42	34.7	546	2 G64803	phosphoglucosylase
4	41	33.9	141	2 T39165	conserved hypotnet
5	41	33.9	157	2 G75266	hypothetical prote
6	41	33.9	402	2 A75054	molybdenum cofacto
7	41	33.9	2044	2 T13704	still life protein
8	41	33.9	2064	2 T13707	still life protein
9	40	33.1	127	2 C69774	transcription regu
10	40	33.1	609	2 T45637	beta-D-glucan exch
11	40	33.1	772	2 T06154	hypothetical prote
12	39	33.2	208	2 D71137	probable transcrip
13	39	33.2	341	2 A40706	extracellular hype
14	39	33.2	688	2 E71845	polynucleotide
15	39	33.2	688	2 E64671	polynucleotide pho
16	38.5	31.8	653	2 T10080	probable carbonate
17	38.5	31.8	1702	2 T14050	protein kinase (EC
18	38	31.4	300	2 E71023	transcription init
19	38	31.4	300	2 E75110	transcription init
20	38	31.4	357	2 UC6703	basic helix-loop-h
21	38	31.4	357	2 I49338	neurogenic differe
22	38	31.4	381	2 A57059	beta-cell E-box tr
23	38	31.4	419	2 S23018	DNA ligase (ATP) (
24	38	31.4	636	2 T45640	beta-D-glucan exch
25	38	31.4	726	2 T20183	hypothetical prote
26	38	31.4	1464	2 T13716	baroque gene prote
27	37	30.6	430	2 S50604	AST2 protein - yea
28	37	30.6	488	2 S40706	hypothetical prote
29	37	30.6	624	2 T04414	probable glucan 1,

30	37	30.6	628	2 T51283	glucan 1,3-beta-gl
31	37	30.6	850	2 T13352	stn-A protein - fr
32	37	30.6	938	2 T05533	hypothetical prote
33	37	30.6	1237	2 A31334	phosphorylase kina
34	37	30.2	1314	2 T09481	matting type silenc
35	36	29.8	85	2 H83389	hypothetical prote
36	36	29.8	115	2 S57269	hypothetical prote
37	36	29.8	153	2 T29164	hypothetical prote
38	36	29.8	160	2 A69849	hypothetical prote
39	36	29.8	238	2 D71189	hypothetical prote
40	36	29.8	268	2 J00961	myb-related protei
41	36	29.8	284	2 UC6198	alpha-tropomyosin
42	36	29.8	303	1 S75782	methanol dehydroge
43	36	29.8	318	2 A43746	nisin resistance p
44	36	29.8	324	2 C42514	H3L protein - vacc
45	36	29.8	324	2 T37369	IMV membrane assoc

## ALIGNMENTS

RESULT 1  
HMGH3Z  
exendin-3 - Mexican beaded lizard  
C:Species: Heloderma horridum (Mexican beaded lizard)  
C:Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 21-Nov-1997  
C:Accession: A23674  
R:Eng, J.; Andrews, P.C.; Kleinman, W.A.; Singh, L.; Raufman, J.P.  
J. Biol. Chem. 265, 20259-20262, 1990.  
A:Title: Purification and structure of exendin-3, a new pancreatic secretagogue isola  
A:Reference number: A23674; MUID:91056067  
A:Accession: A23674  
A:Molecule type: protein  
A:Residues: 1-39 <ENG>  
C:Comment: Exendins are venom components that are thought to bind to receptors for va  
g in secretion of amylase.  
C:Superfamily: glucagon  
C:Keywords: amidated carboxyl end; duplication; secretagogue; venom  
F:39/Modified site: amidated carboxyl end (Ser) #status experimental

Query Match 75.2% Score 91; DB 1; Length 39;  
Best Local Similarity 65.6% Pred. NO. 5.6e-10;  
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 4 GTXXXXXSKQXEEAVRLXXXXLKNKGXSSGA 35  
DB 4 GTFTSLSKQXEEAVRLFIEMLKNKGSSGA 35

RESULT 2  
HMGH3Z  
exendin-4 - Gila monster  
C:Species: Heloderma suspectum (Gila monster)  
C:Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 21-Nov-1997  
C:Accession: A42486  
R:Eng, J.; Kleinman, W.A.; Singh, L.; Singh, G.; Raufman, J.P.  
J. Biol. Chem. 267, 7402-7405, 1992.  
A:Title: Isolation and characterization of exendin-4, an exendin-3 analogue, from Hel  
A:Reference number: A42486; MUID:92218391  
A:Accession: A42486  
A:Molecule type: protein  
A:Residues: 1-39 <ENG>  
C:Comment: Exendin-4 does not stimulate amylase secretion by pancreatic acinar cells.  
C:Superfamily: glucagon  
C:Keywords: amidated carboxyl end; duplication; venom  
F:39/Modified site: amidated carboxyl end (Ser) #status experimental

Query Match 75.2% Score 91; DB 1; Length 39;  
Best Local Similarity 65.6% Pred. NO. 5.6e-10;  
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

```
QY      4 GTXXXXXSKQEEEA VRLXXXXLKGNGKSSGA    35
          ||   ||| ||||| |   ||||| ||||
Db      4 GTFTSDLSKQMEFEA VRLEIWLKNGGPSSGA    35
```

### RESULT 3

Phosphoglucumutase (EC 5.4.2.2) - *Escherichia coli*  
C:Species: *Escherichia coli*  
C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 17-Mar-2000  
C:Accession: G64803; 155076  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co  
A.; Rose, D.J.; Mau, B.; Shaio, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of *Escherichia coli* K-12.  
A:Reference number: A64720; MUID:9742617  
A:Accession: G64803  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-546 <GBAT>  
A:Cross-references: GB:AB000172; GB:U00096; NID:g1786896; PIDN:AACT3782.1; PID:g1786904;  
A:Experimental source: strain K-12, substrain M61655  
R:Lu, M.; Kleckner, N  
J. Bacteriol. 176, 5847-5851, 1994  
A:Title: Molecular cloning and characterization of the pgm gene encoding phosphoglucumut  
A:Reference number: 155076; MUID:94364967  
A:Accession: 155076  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-21, 'M', 23-546 <RES>  
A:Cross-references: EMBL:U00369; NID:g473887; PIDN:AA57067.1; PID:g473888  
C:Genetics:  
A:Gene: pgm  
C:Function:  
A:Description: conversion of D-glucose 1-phosphate into D-glucose 6-phosphate; participate  
C:Superfamily: phosphoglucumutase  
C:Keywords: intramolecular transferase; isomerase; phosphoprotein  
P145/Active site: Ser (phosphoserine intermediate) #status predicted

Query Match  
Post 100%

```

Best Local Similarity 52.9%; Pred. No. 8.9;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0
0y 12 KXKEEAVRLXXXXLKN 28
    |||::|||::|||
Db 529 KQIEKEAVEIVSEVLKN 545

```

RESULT 4

conserved hypothetical protein SPAC8E11.11 - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: J39165  
R:McLean, J.; Harris, D.; Wood, V.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, February 1998  
A:Reference number: Z21831  
A:Accession: J39165  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-141 <MCL>  
A:Cross-references: EMBL,AL021817; PIDB:CA940198.1; GSPDB:GN00066; SPDB:SPAC8E11.11  
A:Experimental source: strain 972h-; cosmid c8E11  
C:Genetics:  
A:Gene: SPDB:SPAC8E11.11  
A:Map position: 1

Query Match	33.98;	Score 41;	DB 2;	Length 141;
Best Local Similarity	34.88;	Pred. No. 3.2;		
Matches	8;	Conservative	5;	Mismatches 10;
				Indels 0;
				Gaps 0

```
QY 12 KQEEEA VRLXXXXLKN GXS SG 34
    | : | : | : | : | : | : | : |
Db 64 KETEVA IEVTKWILSNGVWNG 86
```

RESULT	5
075066	

hypothetical protein DR2500 - *Deinococcus radiodurans* (strain R1)  
C:Species: *Deinococcus radiodurans*  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
C:Accession: G75266  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; Shen, M.; Vamathevan, J.C.; Lam, P.; McDonald, L.; Uitterback, T.; Zalewski, C.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
S: Mille: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.  
A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.  
A:Reference number: A75250; MUID:20036896  
A:Accession: G75266  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-157 <WHI>  
A:Cross-references: GB:AE002079; GB:AE000513; NID:g6460315; PIDN:AEF12045.1; PID:g646  
A:Experimental source: strain R1  
A:Genetics:  
A:Gene: DR2500  
A:Map position: 1  
C:Superfamily: *Deinococcus radiodurans* hypothetical protein DR2500

Query Match

Query Match	33.9%	Score 41;	DB 2;	Length 157;
Best Local Similarity	42.1%;	Pred. No. 3.6;		
Matches	8;	Conservative	4;	Mismatches 7;
				Indels 0;
				Gaps 0;

Db 74 DE

Db 74 DDAVQVEFYRALKNAGLDSG 92

RESULT 6

molybdenum cofactor biosynthesis protein (moea-1) PAB1436 - *Pyrococcus abyssi* (strain  
 C:Species: *Pyrococcus abyssi*  
 C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
 C:Accession: A75054  
 R:anonymous, Genoscope  
 submitted to the EMBL Data Library, July 1999  
 A:Description: *Pyrococcus abyssi* genome sequence: insights into archaeal chromosome s  
 A:Reference number: A75001  
 A:Accession: A75054  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-402 (KMW)  
 A:Cross-references: GB:AU246287; GB:A096836; NID:g5458657; PIDN:CAB50326.1; PID:g545  
 A:Experimental source: strain Orsay  
 C:Genetics:  
 A:Gene: PAB1436  
 C:Superfamily: molybdenum cofactor biosynthesis protein moea-2

Query Match

Query Match	33.9%	Score 41;	DB 2;	Length 402;
Best Local Similarity	39.1%	Pred. No. 9.8;		
Matches	9;	Conservative	4;	Mismatches 10;
			Indels	0;
			Gaps	0;

2  
3  
4  
5

DD 237 KELLEGRAVADLV15GGASGG 232

RESULT 7

T13704  
still life protein type 2 - fruit fly (*Drosophila melanogaster*)  
C.Species: *Drosophila melanogaster*  
C.Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 21-Jul-2000



C:Accession: T13704  
 R:Some, M.; Hoshino, M.; Suzuki, E.; Kuroda, S.; Kambuchi, K.; Nakagoshi, H.; Saigo, K.;  
 Science 275, 543-547, 1997  
 A:Title: Still life, a protein in synaptic terminals of *Drosophila* homologous to GDP-GTF  
 A:Reference number: Z17701; MUID:97153054  
 A:Accession: T13704  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-2044 <SON>  
 A:Cross-references: EMBL:D86546; NID:g1813375; PIDN:BAAL3108.1; PID:g1813376

Query Match 33.9%; Score 41; DB 2; Length 2044;  
 Best Local Similarity 41.7%; Pred. No. 55;  
 Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

OY 12 KXEEAVRLXXXXLKGXSSGA 35

DB 1761 RQIREVNRMSIPMKNEGSSGS 1784

RESULT 8  
 T13707  
 Still life protein type 1 - fruit fly (*Drosophila melanogaster*)

C:Species: *Drosophila melanogaster*  
 C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 21-Jul-2000  
 C:Accession: T13707

R:Some, M.; Hoshino, M.; Suzuki, E.; Kuroda, S.; Kambuchi, K.; Nakagoshi, H.; Saigo, K.;  
 Science 275, 543-547, 1997  
 A:Title: Still life, a protein in synaptic terminals of *Drosophila* homologous to GDP-GTF

A:Reference number: Z17701; MUID:97153054  
 A:Accession: T13707

A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA

A:Residues: 1-2064 <SON>

A:Cross-references: EMBL:D86547; NID:g1813377; PIDN:BAAL3109.1; PID:g1813378

Query Match 33.9%; Score 41; DB 2; Length 2064;  
 Best Local Similarity 41.7%; Pred. No. 55;  
 Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

OY 12 KXEEAVRLXXXXLKGXSSGA 35

DB 1781 RQIREVNRMSIPMKNEGSSGS 1804

RESULT 9  
 C69774  
 transcription regulator phage-related homolog ydcN - *Bacillus subtilis*

C:Species: *Bacillus subtilis*  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000  
 C:Accession: C69774

R:Kunst, F.; Ogatawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berton  
 C.; Bron, S.; Brouillet, S.; Brisch, C.V.; Caldwell, B.; Capiano, V.; Carter, N.M.; Cho  
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier  
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.

Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinis,  
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel  
 Y. M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle

Rieger, M.; Rivolta, C.; Roeha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon  
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot  
 A.; Tamakoshi, A.; Tanaka, T.; Tepstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.;

Winthers, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, K  
 A:Authors: Yoshikawa, H.F.; Zumbach, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
 A:Reference number: A69580; MUID:98044033

A:Accession: C69774  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA  
 A:Residues: 1-127 <KUN>

A:Cross-references: GB:Z99106; GB:AL009126; NID:g2632653; PIDN:CAB12289.1; PID:g26327  
 A:Experimental source: strain 168  
 C:Genetics:  
 A:Gene: ydcN  
 C:Superfamily: probable transcription repressor yowR

Query Match 33.1%; Score 40; DB 2; Length 127;  
 Best Local Similarity 47.1%; Pred. No. 4.4;  
 Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 13 QXEEAVRLXXXXLKG 29

DB 100 EFDEETARLVKKALNG 116

RESULT 10  
 T45637  
 beta-D-glucan exohydrolase-like protein - *Arabidopsis thaliana*

N:Alternate names: protein F13112.60  
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 21-Jul-2000

C:Accession: T45637  
 R:Choine, N.; Robert, C.; Brottier, P.; Winkler, P.; Catolico, L.; Artiguenave, F.;  
 submitted to the Protein Sequence Database, November 1999

A:Reference number: Z23010  
 A:Accession: T45637

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-609 <CHO>  
 A:Cross-references: EMBL:AL133292

A:Experimental source: cultivar Columbia; BAC clone F13112  
 C:Genetics:

A:Map position: 3  
 A:Introns: 57/1; 125/2; 155/3; 204/2; 285/3; 320/3; 381/3; 449/1  
 A:Note: F13112.60

C:Superfamily: beta-glucosidase

Query Match 33.1%; Score 40; DB 2; Length 609;  
 Best Local Similarity 38.5%; Pred. No. 23;  
 Matches 10; Conservative 2; Mismatches 14; Indels 0; Gaps 0;

OY 4 GTXXXXXKXEEAVRLXXXXLKG 29

DB 376 GTGCKREHREAVRKRSLVLRNG 401

RESULT 11  
 T06154  
 hypothetical protein F24J7.162 - *Arabidopsis thaliana*

C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 22-Oct-1999  
 C:Accession: T06154

R:Byran, M.; Vitale, D.; Liqiori, R.; Argirou, A.; De Simone, V.; Bancroft, I.; Mewe  
 submitted to the Protein Sequence Database, April 1999

A:Reference number: Z15493  
 A:Accession: T06154

A:Molecule type: DNA

A:Residues: 1-772 <BEV>  
 A:Cross-references: EMBL:AL021768; GSPDB:GN00062; ATSP:F24J7.162

A:Experimental source: cultivar Columbia; BAC clone F24J7  
 C:Genetics:

A:Gene: ATSP:F24J7.162

A:Map position: 4

A:Introns: 4/2; 42/3; 273/2; 303/2; 342/3; 346/1; 463/3; 485/2; 536/3; 548/3; 576/3;  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown

Query Match 33.1%; Score 40; DB 2; Length 772;  
 Best Local Similarity 37.5%; Pred. No. 30;  
 Matches 9; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

OY 11 SKXEEAVRLXXXXLKGXSSG 34

Db 60 SHKKEEAKKSSSEGLKDNKAGG 83

## RESULT 12

D71137

probable transcription initiation factor IIB - *Pyrococcus horikoshii*C:Species: *Pyrococcus horikoshii*

C:Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 21-Jul-2000

C:Accession: D71137

R:Kawarayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekin

M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi

DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic

A:Reference number: A71000; MUID:98344137

A:Accession: D71137

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-208 &lt;RAW&gt;

A:Cross-references: GB:AP000003; NID:g3236130; PIDN:BAA29958.1; PID:g3257275

A:Experimental source: strain OT3

A:Note: this accession replaces an interim accession for a sequence replaced by GenBank

C:Genetics:

A:Gene: PH0864

C:Superfamily: transcription initiation factor IIB; transcription initiation factor IIB

C:Keywords: transcription initiation

Query Match 32.2%; Score 39; DB 2; Length 208;  
 Best Local Similarity 44.4%; Pred. No. 11;  
 Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 12 SKXEEAVRLXXXXLKNKNG 29

-Db 38 KHVREAVRLYRLKNSG 55

## RESULT 13

A40706

extra-cellular hypersensitive necrosis response elicitor, 34.7K - *Pseudomonas syringae*C:Species: *Pseudomonas syringae*

C:Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 08-Oct-1999

C:Accession: A40706

R:He, S.Y.; Huang, H.C.; Collier, A.

Cell 73, 1255-1266, 1993

A:Title: *Pseudomonas syringae* pv. *syringae* harpin: a protein that is secreted via the

A:Reference number: A40706; MUID:93313957

A:Accession: A40706

A:Status: preliminary

A:Molecule type: protein

A:Residues: 1-341 &lt;HELS&gt;

A:Cross-references: GB:J14775; NID:g349793; PIDN:AAA25839.1; PID:g349794

A:Experimental source: *syringae* 61

A:Note: sequence extracted from NCBI backbone (NCBIN:135085, NCBIPI:135086)

Query Match 32.2%; Score 39; DB 2; Length 341;  
 Best Local Similarity 32.0%; Pred. No. 19;  
 Matches 8; Conservative 4; Mismatches 13; Indels 0; Gaps 0;

QY 5 TXXXXSKXKEEAVRLXXXXLKNKNG 29

Db 27 TTGSTSKALQEVYVKLAELMRNG 51

## RESULT 14

E71845

polyribonucleotide nucleotidyltransferase - *Helicobacter pylori* (strain J99)C:Species: *Helicobacter pylori*

A:Variety: strain J99

C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 24-Sep-1999

C:Accession: E71845

R:Alm, R.A.; Ling, L.S.L.; Molt, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.

Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.  
 Nature 397, 176-180, 1999

A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric p

A:Reference number: A71800; MUID:99120557

A:Accession: E71845

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-688 <ARN>

A:Cross-references: GB:AE001541; GB:AE001439; NID:g4155724; PIDN:AMD06718.1; PID:g415

A:Experimental source: strain J99

C:Genetics:

A:Gene: pnp

C:Superfamily: polyribonucleotide nucleotidyltransferase alpha chain

Query Match 32.2%; Score 39; DB 2; Length 688;  
 Best Local Similarity 36.0%; Pred. No. 40;  
 Matches 9; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 11 SKXEEAVRLXXXXLKNKNGSSGA 35

Db 415 SIKNKEQVIRLVSEILNNGSSSMA 439

## RESULT 15

E64671

polynucleotide phosphorylase - *Helicobacter pylori* (strain 26695)C:Species: *Helicobacter pylori*

C:Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 24-Sep-1999

C:Accession: E64671

R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.

Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McK

son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watney,

Nature 388, 539-547, 1997

A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karp, P.D.; Smith, H.O.; Fraser,

A:Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.

A:Reference number: A64520; MUID:97394467

A:Accession: E64671

A:Status: preliminary

A:Molecule type: nucleic acid sequence not shown; translation not shown

A:Residues: 1-688 &lt;TON&gt;

A:Cross-references: GB:AE000627; GB:AE000511; NID:g2314373; PIDN:AA008258.1; PID:g231

C:Superfamily: polyribonucleotide nucleotidyltransferase alpha chain

Query Match 32.2%; Score 39; DB 2; Length 688;  
 Best Local Similarity 36.0%; Pred. No. 40;  
 Matches 9; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 11 SKXEEAVRLXXXXLKNKNGSSGA 35

Db 415 SIKNKEQVIRLVSEILNNGSSSMA 439

Search completed: February 13, 2001, 16:08:13  
 Job time: 128 sec

Page 5 of 5

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2001, 16:06:05 ; Search time 22.07 Seconds  
(without alignments)  
32.546 Million cell updates/sec

Title: US-08-908-867-38

Perfect score: 121  
Sequence: 1 XXXGTXXXXXXKQXEEAVRLXXXXLKNKGXSGAXXXXX 40

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6PCTUS.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	91	75.2	39 1	US-08-066-480-1
2	91	75.2	39 1	US-08-066-480-2
3	83	68.6	31 1	US-08-066-480-5
4	75	62.0	31 1	US-08-066-480-3
5	62.0	31	1	US-08-066-480-4
6	41	33.9	589 2	US-08-317-305-2
7	41	33.9	589 2	US-08-317-305-4
8	41	33.9	589 3	US-08-862-508-2
9	41	33.9	589 3	US-08-862-508-4
10	41	33.9	589 4	PCT-US95-12508-2
11	41	33.9	589 4	PCT-US95-12508-4
12	39	32.2	341 1	US-08-062-024B-5
13	39	32.2	341 1	US-08-891-254-5
14	39	32.2	341 2	US-08-756-407-5
15	39	32.2	341 2	US-08-819-539-5
16	39	32.2	341 2	US-09-030-270A-5
17	39	32.2	341 4	PCT-US94-05014-5
18	39	32.2	341 4	PCT-US96-08819-5
19	39	31.8	651 2	US-08-492-027A-1
20	38.5	31.8	655 2	US-08-492-027A-6
21	38	31.4	357 1	US-08-552-142A-2
22	38	31.4	357 1	US-08-910-973-2
23	38	31.4	357 4	PCT-US95-05741-2
24	36	29.8	318 1	US-08-220-958-4
25	35	28.9	156 1	US-08-552-142A-9
26	35	28.9	156 1	US-08-910-973-9
27	35	28.9	156 4	PCT-US95-05741-9
28	35	28.9	156 4	US-08-552-142A-15

29	35	28.9	356 1 US-08-910-973-15	Sequence 15, Appli
30	35	28.9	356 3 US-09-234-332-7	Sequence 7, Appli
31	35	28.9	356 3 US-09-234-332-8	Sequence 8, Appli
32	35	28.9	546 2 US-08-492-027A-8	Sequence 8, Appli
33	35	28.9	619 3 US-08-813-150-6	Sequence 6, Appli
34	35	28.9	777 2 US-08-231-193A-16	Sequence 16, Appli
35	35	28.9	777 2 US-08-486-273A-16	Sequence 16, Appli
36	35	28.9	777 3 US-08-480-474-16	Sequence 16, Appli
37	35	28.9	777 3 US-08-940-086A-16	Sequence 16, Appli
38	35	28.9	854 2 US-08-231-193A-32	Sequence 32, Appli
39	35	28.9	854 2 US-08-486-273A-32	Sequence 32, Appli
40	35	28.9	854 3 US-08-480-474-32	Sequence 32, Appli
41	35	28.9	854 3 US-08-940-086A-32	Sequence 32, Appli
42	35	28.9	863 3 US-08-436-332B-2	Sequence 2, Appli
43	35	28.9	870 2 US-08-231-193A-30	Sequence 30, Appli
44	35	28.9	870 2 US-08-486-273A-30	Sequence 30, Appli
45	35	28.9	870 3 US-08-480-474-30	Sequence 30, Appli

#### ALIGNMENTS

RESULT 1  
US-08-066-480-1  
; Sequence 1, Application US/08066480  
; Patent No. 5424286  
; GENERAL INFORMATION:  
; APPLICANT: Eng, John  
; TITLE OF INVENTION: Pharmaceutical Compositions And Use of  
; TITLE OF INVENTION: Exendin-3 and Exendin-4 for Treatment of Diabetes Mellitus  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Allogreffe & Witcoff, Ltd.  
; STREET: 10 S. Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/066,480  
; FILING DATE: 24-MAR-1993  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McDonnell, John J  
; REGISTRATION NUMBER: 26,949  
; REFERENCE/DOCKET NUMBER: 93,084  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-715-1000  
; TELEFAX: 312-715-1234  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 39 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FEATURE:  
; NAME/KEY: Peptide  
; LOCATION: 1..39  
; OTHER INFORMATION: /label= Exendin-3  
; US-08-066-480-1

Query Match 75.2% Score 91; DB 1; Length 39;  
Best Local Similarity 65.6% Pred No. 5.9e-10;  
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
QY 4 GTXXXXXKQXEEAVRLXXXXLKNKGXSGA 35

Db 4 GTFTSDLSKQMEEAVALFLEWLKNGPSSGA 35

RESULT 2  
US-08-066-480-2

; Sequence 2, Application US/08066480  
; Patent No. 5424286  
; GENERAL INFORMATION:  
; APPLICANT: Eng, John  
; TITLE OF INVENTION: Pharmaceutical Compositions And Use of  
; TITLE OF INVENTION: Exendin-3 and Exendin-4 for Treatment of Diabetes Mellitus  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Allegretti & Witcoff, Ltd.  
; STREET: 10 S. Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/066,480  
; FILING DATE: 24-MAR-1993  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McDonnell, John J  
; REGISTRATION NUMBER: 26,949  
; REFERENCE/DOCKET NUMBER: 93,084  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-715-1000  
; TELEFAX: 312-715-1234  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 39 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FEATURE:  
; NAME/KEY: Peptide  
; LOCATION: 1..39  
; OTHER INFORMATION: /label= Exendin-4  
US-08-066-480-2

Query Match 75.2%; Score 91; DB 1; Length 39;  
Best Local Similarity 65.6%; Pred. No. 5.9e-10;  
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 4 GTXXXXXKQEEAVRLXXXXLKNGXSSGA 35  
Db 4 GTFTSDLSKQMEEAVALFLEWLKNGPSSGA 35

RESULT 3  
US-08-066-480-5  
; Sequence 5, Application US/08066480  
; Patent No. 5424286  
; GENERAL INFORMATION:  
; APPLICANT: Eng, John  
; TITLE OF INVENTION: Pharmaceutical Compositions And Use of  
; TITLE OF INVENTION: Exendin-3 and Exendin-4 for Treatment of Diabetes Mellitus  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Allegretti & Witcoff, Ltd.  
; STREET: 10 S. Wacker Drive  
; CITY: Chicago  
; STATE: Illinois

COUNTRY: USA.

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/066,480

FILING DATE: 24-MAR-1993

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: McDonnell, John J

REGISTRATION NUMBER: 26,949

REFERENCE/DOCKET NUMBER: 93,084

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-715-1000

TELEFAX: 312-715-1234

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 31 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Peptide

LOCATION: 1..31

OTHER INFORMATION: /note= "Exendin-4(9-39)"

US-08-066-480-5

Query Match 68.6%; Score 83; DB 1; Length 31;  
Best Local Similarity 76.0%; Pred. No. 1.3e-08;  
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 11 SKQEEAAVRLXXXXLKNGXSSGA 35  
Db 3 SKQEEAAVRLFLEWLKNGPSSGA 27

RESULT 4  
US-08-066-480-3  
; Sequence 3, Application US/08066480  
; Patent No. 5424286  
; GENERAL INFORMATION:

APPLICANT: Eng, John

TITLE OF INVENTION: Pharmaceutical Compositions And Use of

TITLE OF INVENTION: Exendin-3 and Exendin-4 for Treatment of Diabetes Mellitus

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: Allegretti & Witcoff, Ltd.

STREET: 10 S. Wacker Drive

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/066,480

FILING DATE: 24-MAR-1993

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: McDonnell, John J

REGISTRATION NUMBER: 26,949

REFERENCE/DOCKET NUMBER: 93,084

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-715-1000

TELEFAX: 312-715-1234  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 31 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..31  
OTHER INFORMATION: /label= Exendin-1-31  
OTHER INFORMATION: /note= "Exendin-4(1-31)"  
US-08-066-480-3

Query Match 62.0%; Score 75; DB 1; Length 31;  
Best Local Similarity 63.0%; Pred. No. 3.4e-07;  
Matches 17; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 4 GTXXXXXSKQXEEAVRLXXXXLKNKG 30  
DB 4 GTFTSDLSKQMEEAVALFTEWLKNG 30

RESULT 5  
US-08-066-480-4  
Sequence 4, Application US/08066480  
Patent No. 5424286  
GENERAL INFORMATION:  
APPLICANT: Eng, John  
TITLE OF INVENTION: Pharmaceutical Compositions And Use of  
NUMBER OF INVENTION: Exendin-3 and Exendin-4 for Treatment of Diabetes Mellitus  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Allegretti & Wilcoff, Ltd.  
STREET: 10 S. Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/066,480  
FILING DATE: 24-MAR-1993  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: McDonnell, John J  
REGISTRATION NUMBER: 26,949  
REFERENCE/DOCKET NUMBER: 93,084  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-715-1000  
TELEFAX: 312-715-1234  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 31 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..31  
OTHER INFORMATION: /label= Y31-Exendin4  
OTHER INFORMATION: /note= "Y-31-Exendin-4(1-31)"  
US-08-066-480-4

Query Match 62.0%; Score 75; DB 1; Length 31;

Best Local Similarity 63.0%; Pred. No. 3.4e-07;  
Matches 17; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
QY 4 GTXXXXXSKQXEEAVRLXXXXLKNKG 30  
DB 4 GTFTSDLSKQMEEAVALFTEWLKNG 30

RESULT 6  
US-08-317-305-2  
Sequence 2, Application US/08317305  
Patent No. 5863744  
GENERAL INFORMATION:  
APPLICANT: Avraham, Shalom  
APPLICANT: Avraham, Hava  
APPLICANT: Groopman, Jerome E.  
TITLE OF INVENTION: NOVEL PROTEIN MARKER AND DNA ENCODING  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes  
STREET: Ten Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/317,305  
FILING DATE: 03-OCT-1994  
CLASSIFICATION: 436  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Ph.D., Kathleen A.  
REGISTRATION NUMBER: 34,380  
REFERENCE/DOCKET NUMBER: DH-001XX  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-2290  
TELEFAX: (617) 451-0313  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 589 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-317-305-2

Query Match 33.9%; Score 41; DB 2; Length 589;  
Best Local Similarity 45.0%; Pred. No. 12;  
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 11 SKQXEEAVRLXXXXLKNKG 30  
DB 262 SKEIVEAIRCKLQNDG 281

RESULT 7  
US-08-317-305-4  
Sequence 4, Application US/08317305  
Patent No. 5863744  
GENERAL INFORMATION:  
APPLICANT: Avraham, Shalom  
APPLICANT: Avraham, Hava  
APPLICANT: Groopman, Jerome E.  
TITLE OF INVENTION: NOVEL PROTEIN MARKER AND DNA ENCODING  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes

STREET: Ten Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA: US/08/317,305  
APPLICATION NUMBER: US/08/317,305  
FILING DATE: 03-OCT-1994  
CLASSIFICATION: 436  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Ph.D., Kathleen A.  
REGISTRATION NUMBER: 34,380  
REFERENCE/DOCKET NUMBER: DH-001XX  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-2290  
TELEFAX: (617) 451-0313  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 589 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-317-305-4

Query Match 33.9%; Score 41; DB 2; Length 589;  
Best Local Similarity 45.0%; Pred. No. 12;  
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

.QY 11 SKOXEEAVRLXXXXKNG 30  
||: |||:| | :| |  
Db 262 SKEIVEAIRCKIKLQNDG 281

RESULT 8  
US-08-862-508-2  
Sequence 2, Application US/08862508  
Patent No. 6066451  
GENERAL INFORMATION:  
APPLICANT: Avraham, Shalom  
APPLICANT: Avraham, Hava  
APPLICANT: Groopman, Jerome E.  
TITLE OF INVENTION: NOVEL NEURAL CELL PROTEIN MARKER RR/B AND DNA  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lahive & Cockfield, LLP  
STREET: 28 State Street  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA: US/08/862,508  
APPLICATION NUMBER: US/08/862,508  
FILING DATE: 23-MAY-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/317,305  
FILING DATE: 03-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Silverl, Jean M.  
REGISTRATION NUMBER: 39,030  
REFERENCE/DOCKET NUMBER: NER-259DV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400

TELEFAX: (617) 742-4214  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 589 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-862-508-2

Query Match 33.9%; Score 41; DB 3; Length 589;  
Best Local Similarity 45.0%; Pred. No. 12;  
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

.QY 11 SKOXEEAVRLXXXXKNG 30  
||: |||:| | :| |  
Db 262 SKEIVEAIRCKIKLQNDG 281

RESULT 9  
US-08-862-508-4  
Sequence 4, Application US/08862508  
Patent No. 6066451  
GENERAL INFORMATION:  
APPLICANT: Avraham, Shalom  
APPLICANT: Avraham, Hava  
APPLICANT: Groopman, Jerome E.  
TITLE OF INVENTION: NOVEL NEURAL CELL PROTEIN MARKER RR/B AND DNA  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lahive & Cockfield, LLP  
STREET: 28 State Street  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA: US/08/862,508  
APPLICATION NUMBER: US/08/862,508  
FILING DATE: 23-MAY-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/317,305  
FILING DATE: 03-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Silverl, Jean M.  
REGISTRATION NUMBER: 39,030  
REFERENCE/DOCKET NUMBER: NER-259DV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 742-4214  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 589 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-862-508-4

Query Match 33.9%; Score 41; DB 3; Length 589;  
Best Local Similarity 45.0%; Pred. No. 12;  
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

.QY 11 SKOXEEAVRLXXXXKNG 30  
||: |||:| | :| |  
Db 262 SKEIVEAIRCKIKLQNDG 281

RESULT 10



PCT-US95-12508-2  
; Sequence 2, Application PC/TUS9512508  
; GENERAL INFORMATION:  
; APPLICANT: Avraham, Shalom  
; APPLICANT: Avraham, Hava  
; APPLICANT: Grozman, Jerome E.  
; TITLE OF INVENTION: NOVEL PROTEIN MARKER AND DNA ENCODING SAME  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lahive & Cockfield  
; STREET: 60 State Street, Suite 510  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/12508  
; FILING DATE: 29-SEP-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/317,305  
; FILING DATE: 03-OCT-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Silverl, Jean M.  
; REGISTRATION NUMBER: 39,030  
; REFERENCE/DOCKET NUMBER: NER-259PC  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 227-7400  
; TELEFAX: (617) 227-5941  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 589 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; PCT-US95-12508-2

Query Match 33.9%; Score 41; DB 4; Length 589;  
Best Local Similarity 45.0%; Pred. No. 12;  
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 11 SKQEEEAIVRLXXXXKNG 30  
DB 262 SKEIVEAIFCKRLKILNDG 281

RESULT 11  
PCT-US95-12508-4  
; Sequence 4, Application PC/TUS9512508  
; GENERAL INFORMATION:  
; APPLICANT: Avraham, Shalom  
; APPLICANT: Avraham, Hava  
; APPLICANT: Grozman, Jerome E.  
; TITLE OF INVENTION: NOVEL PROTEIN MARKER AND DNA ENCODING SAME  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lahive & Cockfield  
; STREET: 60 State Street, Suite 510  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/12508  
; FILING DATE: 29-SEP-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/317,305  
; FILING DATE: 03-OCT-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Silverl, Jean M.  
; REGISTRATION NUMBER: 39,030  
; REFERENCE/DOCKET NUMBER: NER-259PC  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 227-7400  
; TELEFAX: (617) 227-5941  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 589 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; PCT-US95-12508-4

Query Match 33.9%; Score 41; DB 4; Length 589;  
Best Local Similarity 45.0%; Pred. No. 12;  
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 11 SKQEEEAIVRLXXXXKNG 30  
DB 262 SKEIVEAIFCKRLKILNDG 281

RESULT 12  
US-08-062-024B-5  
; Sequence 5, Application US/08062024B  
; Patent No. 5708139  
; GENERAL INFORMATION:  
; APPLICANT: Alan Collmer and Sheng-Yang He  
; TITLE OF INVENTION: Pseudomonas syringae pv. syringae hrpZ Gene  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Yahwak & Associates  
; STREET: 25 Skytop Drive  
; CITY: Trumbull  
; STATE: Connecticut  
; COUNTRY: USA  
; ZIP: 06611  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: floppy disk  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: Microsoft word 4.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/062,024B  
; FILING DATE: May 17th 1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: George M. Yahwak  
; REGISTRATION NUMBER: 26,824  
; REFERENCE/DOCKET NUMBER: CRF D-1425  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (203)268-1951  
; TELEFAX: (203)268-1951  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 341 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-062-024B-5

Query Match 32.2%; Score 39; DB 1; Length 341;  
Best Local Similarity 32.0%; Pred. No. 15;





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2001, 16:06:04 ; Search time 26.49 Seconds  
(without alignments)  
51.633 Million cell updates/sec

Title: US-08-908-867-38

Perfect score: 121  
Sequence: 1 XXXGTXXXXXXKQXEEAVRLXXXXLKNCGXSGAXXXXX 40

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT:\*  
2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT:\*  
3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT:\*  
4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT:\*  
5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT:\*  
6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT:\*  
7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT:\*  
8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT:\*  
9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT:\*  
10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT:\*  
11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT:\*  
12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT:\*  
13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT:\*  
14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT:\*  
15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT:\*  
16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT:\*  
17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT:\*  
18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT:\*  
19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT:\*  
20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT:\*  
21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	94	77.7	37	20	Y17618
2	93	76.9	37	20	Y17606
3	93	76.9	37	20	Y24869
4	93	76.9	37	20	Y24853
5	93	76.9	37	20	Y24854
6	92	76.0	35	20	Y31535
7	92	76.0	35	20	Y24839
8	92	76.0	35	20	Y17608
9	92	76.0	36	20	Y31533
10	92	76.0	36	20	Y24856
11	92	76.0	36	20	Y24837
12	92	76.0	36	20	Y17619

13	92	76.0	36	20	Y17605	Exendin agonist pe
14	92	76.0	37	20	Y31531	Exendin agonist pe
15	92	76.0	37	20	Y31549	Exendin agonist pe
16	92	76.0	37	20	Y31550	Exendin agonist pe
17	92	76.0	37	20	Y31565	Exendin agonist pe
18	92	76.0	37	20	Y24855	Exendin agonist pe
19	92	76.0	37	20	Y24835	Exendin agonist pe
20	92	76.0	38	20	Y31529	Exendin agonist pe
21	92	76.0	38	20	Y31547	Exendin agonist pe
22	92	76.0	38	20	Y24851	Exendin agonist pe
23	92	76.0	38	20	Y24833	Exendin agonist pe
24	92	76.0	38	20	Y17616	Exendin agonist pe
25	92	76.0	38	20	Y17603	Exendin agonist pe
26	92	76.0	39	19	W61773	Exendin agonist pe
27	92	76.0	39	20	W61763	Exendin agonist pe
28	92	75.2	35	20	Y31534	Exendin agonist pe
29	91	75.2	35	20	Y31553	Exendin agonist pe
30	91	75.2	35	20	Y24857	Exendin agonist pe
31	91	75.2	35	20	Y24838	Exendin agonist pe
32	91	75.2	35	20	Y17620	Exendin agonist pe
33	91	75.2	35	20	Y17607	Exendin agonist pe
34	91	75.2	35	21	Y78961	Exendin-4 (1-35)
35	91	75.2	36	20	Y31532	Exendin agonist pe
36	91	75.2	36	20	Y31552	Exendin agonist pe
37	91	75.2	36	20	Y24836	Exendin agonist pe
38	91	75.2	36	21	Y78960	Exendin-4 (1-36)
39	91	75.2	37	20	Y31530	Exendin agonist pe
40	91	75.2	37	20	Y31551	Exendin agonist pe
41	91	75.2	37	21	Y78959	Exendin-4 (1-37)
42	91	75.2	38	20	Y31528	Exendin agonist pe
43	91	75.2	38	20	Y31548	Exendin agonist pe
44	91	75.2	38	20	Y24852	Exendin agonist pe
45	91	75.2	38	20	Y24832	Exendin agonist pe

#### ALIGNMENTS

RESULT 1	
ID Y17618	standard; peptide; 37 AA.
AC Y17618;	
XX	
DT 09-AUG-1999	(first entry)
DE	Exendin agonist peptide #84.
XX	
KW	Exendin; agonist; Heloderma sp.; Gila monster; venom; lizard;
KW	diabetes mellitus type I; diabetes mellitus type II; hyperglycaemia;
KW	hypoglycaemia; plasma glucose; gastric emptying; stomach emptying.
XX	
OS	Synthetic.
OS	Heloderma sp.
XX	
PN	WO9925728-A1.
XX	
PD 27-MAY-1999.	
XX	
PF 13-NOV-1998;	98WO-US24273.
XX	
PR 14-NOV-1997;	97US-0066029.
XX	
PA (AMYL-) AMYLIN PHARM INC.	
XX	
PI Beeley NRA, Prickekt KS;	
XX	
DR WPI; 1999-347456/29.	
XX	
PT Peptide agonists of exendin - delay stomach emptying, for treating	
PT diabetes and hypo- or hyper-glycaemia	
XX	
PS Claim 28; Fig 4; 144p; English.	



us-08-908-867-38.rag

Sequence 37 AA:

PF 13-NOV-1998; 98WO-US24210.

[illegible]







```

XX DE Extendin agonist peptide #29.
XX XX
XX RM Extendin; agonist; Heloderma sp.; Gila monster; venom; lizard.
XX KW diabetes mellitus type I; diabetes mellitus type II; hyperglycaemia;
XX RW hypoglycaemia; plasma glucose; gastric emptying; stomach emptying.
XX OS Synthetic.
OS Heloderma sp.
XX PM W09925727-A2.
XX PD 27-MAY-1999.
XX PF 13-NOV-1998; 98WO-US24210.
XX PR 14-NOV-1997; 97US-0065442.
XX PA (AMYL-) AMYLIN PHARM INC.
XX PI Beeley NRA, Prickett KS;
XX PS WPI. 1999-394773/33.
XX PT New extendin agonist peptides - can regulate gastric motility and
XX PP slow gastric emptying, used for treating, e.g. diabetes
XX Claim 18; Fig 4; 108pp; English.
CC Y24809 to Y24877 represent extendin agonist peptides which can regulate
CC gastric motility and slow gastric emptying. The peptides can be used for
CC treating e.g. diabetes or hyperglycaemic or hypoglycaemic conditions.
CC The peptides are extendin agonists which have activity as agents to
CC regulate gastric motility and to slow gastric emptying, as evidenced by
CC the ability to reduce post-prandial glucose levels in mammals. They can
CC be used for the treatment of Type I and II diabetes and hyperglycaemic
CC or hypoglycaemic conditions. They can also be used for the treatment of
CC disorders which would be benefited by agents which lower plasma glucose
CC levels and in treatment of disorders which would be benefited with
CC agents useful in delaying and/or slowing gastric emptying.
SQ Sequence 36 AA:

Query Match 76.0%; Score 92; DB 20; Length 36;
Best Local Similarity 65.6%; Pred. No. 2.7e-10;
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0.

OY 4 GTXXXXXSKOXEEAVRLKXXXLKNGGXSSGA 35
   ||| ||||| | ||||| |||
Db 4 gftsdlskqleeeavrlfiefllknngpsssa 35

RESULT 12
YI7619
YI7619 standard; peptide: 36 AA.
XX ID YI7619;
XX AC YI7619;
XX DT 09-AUG-1999 (first entry)
XX DE Extendin agonist peptide #85.
XX KM Extendin; agonist; Heloderma sp.; Gila monster; venom; lizard.
XX KW diabetes mellitus type I; diabetes mellitus type II; hyperglycaemia;
XX RM hypoglycaemia; plasma glucose; gastric emptying; stomach emptying.
XX OS Synthetic.
OS Heloderma sp.
XX PM W09925728-A1.
XX PD 27-MAY-1999.
```

[illegible]

CC peptides that are found in the venom of the Gila-monster, a lizard  
 CC endogenous to Arizona and Northern Mexico. The peptide agonists are  
 CC used to treat diabetes mellitus (types I or II), hyperglycaemia or  
 CC hypoglycaemia. They can also be used for in vitro and in vivo studies  
 CC on extendin and their agonists. They regulate gastric motility and slow  
 CC gastric emptying (resulting in lower post-prandial glucose levels).  
 XX  
 SQ Sequence 36 AA;

Query Match 76.0%; Score 92; DB 20; Length 36;  
 Best Local Similarity 65.6%; Pred. No. 2.7e-10;  
 Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 4 GTXXXXXSKOXEEAVRLXXXXLKNGXSGA 35  
 || ||| ||||| ||||| ||||| |||||  
 DB 4 gtfalslstkqmeeeavrlfiefllkngpssga 35

## RESULT 14

ID Y31531 Y31531 standard; peptide; 37 AA.

AC Y31531;

DT 08-NOV-1999 (first entry)

DE Extendin agonist peptide.

KM Extendin; agonist; GLP-1; glucagon-like peptide; toxic hypervolemia;

KM diuresis; renal plasma flow; glomerular filtration rate; pre-eclampsia;

KM eclampsia of pregnancy; cardiac contractility; renal failure; diuretic;

KM congestive heart failure; nephrotic syndrome; pulmonary edema; cirrhosis;

KM hypertension; urine flow.

XX Synthetic.

OS Heloderma sp.

PN W09940788-A1.

PD 19-AUG-1999.

PF 05-FEB-1999; 99WO-US02554.

PR 13-FEB-1998; 98US-0075122.

PA (AMYL-) AMYLIN PHARM INC.

PI Beeley NRA, Prickett K, Vine W, Young AA;

DR WPI; 1999-527332/44.

PT Increasing urine flow by administering peptides or peptide agonists

XX Example 30; Page 46; 94pp; English.

CC The invention relates to new methods of increasing urine flow that  
 CC comprises administering an extendin or extendin agonist, or a GLP-1  
 CC (glucagon-like peptide) or GLP-1 agonist. The new methods using an  
 CC extendin, extendin agonist, GLP-1 or GLP-1 agonist are useful for  
 CC preventing or alleviating a disorder associated with toxic hypervolemia  
 CC (renal failure, congestive heart failure, nephrotic syndrome, pulmonary  
 CC edema, cirrhosis, or hypertension). They can also be used for inducing  
 CC rapid diuresis, preparing an individual for surgical procedure,  
 CC increasing renal plasma flow and glomerular filtration rate, treating  
 CC pre-eclampsia or eclampsia of pregnancy, and increasing a condition/  
 CC disorder that can be alleviated by increasing cardiac contractility  
 CC (congestive heart failure, pulmonary edema, systemic edema or renal

CC failure). Unlike prior art diuretics, the new methods increase urine  
 CC excretion and sodium excretion without increasing potassium loss, and are  
 CC fast acting. They have a prolonged duration of action, are isotropic,  
 CC have a low toxicity, and are easily administered intravenously.  
 CC Sequences Y31505-560 represent examples of extendin agonists compounds.  
 XX  
 SQ Sequence 37 AA;

Query Match 76.0%; Score 92; DB 20; Length 37;  
 Best Local Similarity 65.6%; Pred. No. 2.8e-10;  
 Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 4 GTXXXXXSKOXEEAVRLXXXXLKNGXSGA 35  
 || ||| ||||| ||||| ||||| |||||  
 DB 4 gtfdsldskqleeeavrlfiefllkngpssga 35

## RESULT 15

ID Y31549 Y31549 standard; peptide; 37 AA.

AC Y31549;

DT 08-NOV-1999 (first entry)

DE Extendin agonist peptide.

KM Extendin; agonist; GLP-1; glucagon-like peptide; toxic hypervolemia;

KM diuresis; renal plasma flow; glomerular filtration rate; pre-eclampsia;

KM eclampsia of pregnancy; cardiac contractility; renal failure; diuretic;

KM congestive heart failure; nephrotic syndrome; pulmonary edema; cirrhosis;

KM hypertension; urine flow.

XX Synthetic.

OS Heloderma sp.

PN W09940788-A1.

PD 19-AUG-1999.

PF 05-FEB-1999; 99WO-US02554.

PR 13-FEB-1998; 98US-0075122.

PA (AMYL-) AMYLIN PHARM INC.

PI Beeley NRA, Prickett K, Vine W, Young AA;

DR WPI; 1999-527332/44.

PT Increasing urine flow by administering peptides or peptide agonists

XX Example 48; Page 55; 94pp; English.

CC The invention relates to new methods of increasing urine flow that  
 CC comprises administering an extendin or extendin agonist, or a GLP-1  
 CC (glucagon-like peptide) or GLP-1 agonist. The new methods using an  
 CC extendin, extendin agonist, GLP-1 or GLP-1 agonist are useful for  
 CC preventing or alleviating a disorder associated with toxic hypervolemia  
 CC (renal failure, congestive heart failure, nephrotic syndrome, pulmonary  
 CC edema, cirrhosis, or hypertension). They can also be used for inducing  
 CC rapid diuresis, preparing an individual for surgical procedure,  
 CC increasing renal plasma flow and glomerular filtration rate, treating  
 CC pre-eclampsia or eclampsia of pregnancy, and increasing a condition/  
 CC disorder that can be alleviated by increasing cardiac contractility

CC (congestive heart failure, pulmonary edema, systemic edema or renal  
CC failure). Unlike prior art diuretics, the new methods increase urine  
CC excretion and sodium excretion without increasing potassium loss, and are  
CC fast acting. They have a prolonged duration of action, are inotropic,  
CC have a low toxicity, and are easily administered intravenously.  
CC Sequences Y31505-560 represent examples of exendin agonists compounds.  
XX  
SQ Sequence 37 AA;

Query Match 76.0%; Score 92; DB 20; Length 37;  
Best Local Similarity 68.8%; Pred. No. 2.8e-10;  
Matches 22; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 4 GTXXXXXSKOXEEAVRLXXXXLKNKGXSSGA 35  
|| ||||| |||||  
Db 4 gftsdlskqmeeeavrllfewlknngxssga 35

Search completed: February 13, 2001, 16:06:38  
/Job time: 34 sec